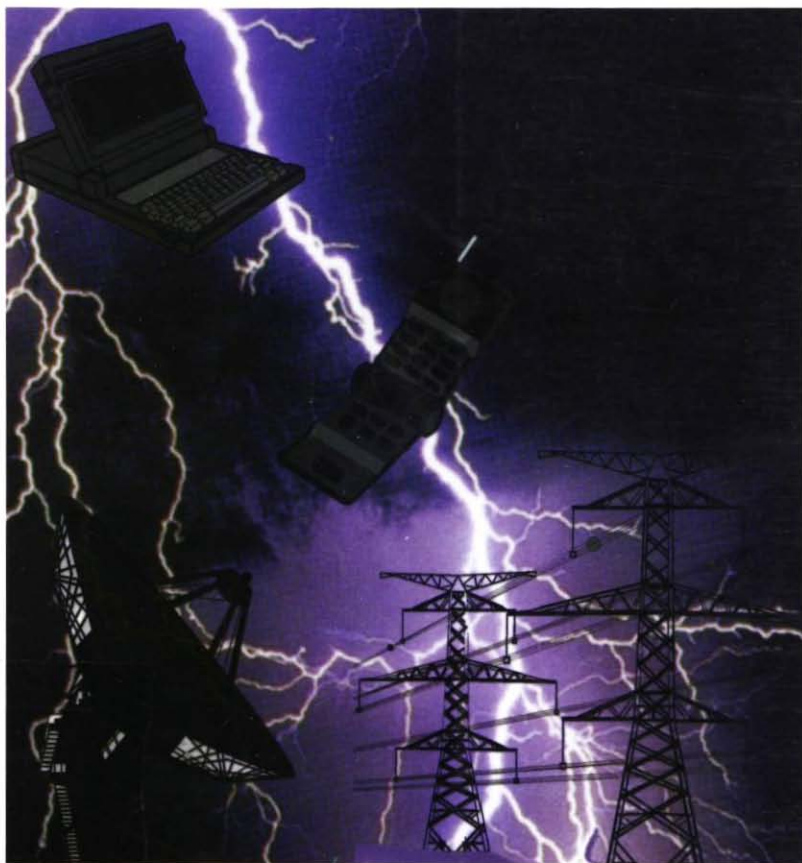


Principios de comunicaciones digitales

José Guadalupe Viveros Talavera





JOSÉ GUADALUPE VIVEROS TALAVERA cursó la Licenciatura en Física en la Facultad de Ciencias de la UNAM; la Maestría en Ciencias, opción Instrumentación Electrónica y Telecomunicaciones, en el Centro de Investigación Científica y de Estudios Superiores de Ensenada, B.C. Ha laborado en diversas instituciones de docencia e investigación, así como en empresas privadas de telecomunicaciones. Actualmente se desempeña como profesor-investigador en el Departamento de Electrónica de la UAM-Azcapotzalco. Sus principales áreas de trabajo son los sistemas de telefonía fija y celular.

PRINCIPIOS DE COMUNICACIONES DIGITALES

COLECCIÓN

Libros de Texto y Manuales de Práctica

SERIE

Material de apoyo a la docencia

(Teoría y prácticas de laboratorio, problemarios)

#217939
P.B. 2893843

Principios de comunicaciones digitales

José Guadalupe Viveros Talavera



UNIVERSIDAD AUTÓNOMA METROPOLITANA

Rector General

José Luis Gázquez Mateos

Secretario General

Edmundo Jacobo Molina

UNIDAD AZCAPOTZALCO

Rectora

Mtra. Mónica de la Garza Malo

Secretario

Lic. Guillermo Ejea Mendoza

Coordinador de Extensión Universitaria

Lic. Enrique López Aguilar

Jefa de la Sección de Producción y Distribución Editoriales

Lic. Silvia Aboytes Perete

Portada

Virginia Flores/Sans Serif Editores

Composición tipográfica, diseño, producción y cuidado editorial

Sans Serif Editores, telfax 5674 60 91

Primera edición 1999

ISBN: 970-654-242-6

© Universidad Autónoma Metropolitana
Unidad Azcapotzalco
Av. San Pablo núm. 180
México, 02200, D.F.

Impreso en México

Printed in Mexico

CAPÍTULO I

INTRODUCCIÓN

I.1. ALGUNAS DEFINICIONES

SE ENTIENDE POR COMUNICACIÓN el proceso por el cual la información se transfiere desde un punto en el espacio y tiempo llamado fuente a otro punto denominado destinatario. En consecuencia, un sistema de comunicación es la totalidad de mecanismos que proveen el enlace de la información entre la fuente y el destinatario. Un sistema de comunicaciones eléctrico efectúa esta función primariamente, pero no en forma exclusiva, mediante el uso de dispositivos eléctricos.

Todos los sistemas de comunicación tienen la misma función básica, la transmisión de la información, y nosotros estamos particularmente interesados en la transmisión de la información en forma eléctrica. Haciendo de lado por un momento el concepto de información, el que analizaremos en detalle en el capítulo correspondiente, nos centraremos en la manifestación física de la información tal como la produce la fuente, es decir, en el mensaje.

Si bien las fuentes de información son diversas, incluyendo tanto máquinas como seres humanos, resulta posible y conveniente distinguir dos tipos diferentes de mensajes: analógicos y digitales.

Un mensaje analógico es una magnitud física, variable en el tiempo según una función continua. Es decir, dentro de cierto rango puede tomar infinitos valores. Así tenemos por ejemplo la variación de presión sonora que produce la música o la palabra, la intensidad de luz en una escena, el nivel de combustible en un depósito, etcétera.

Un mensaje digital es una secuencia ordenada de símbolos, seleccionados de un conjunto finito de elementos discretos. Ejemplos de mensajes digitales son las letras de un texto, los orificios de una cinta o una tarjeta perforada, los puntos y rayas del código Morse, etcétera.

Ya sea analógico o digital, el mensaje que produce la fuente no es eléctrico y será necesario un dispositivo que se denomina transductor de entrada, para convertirlo en una magnitud eléctrica. Dicha magnitud variable en el tiempo, ya sea una corriente o una tensión, que representa el mensaje en forma eléctrica, se denomina señal. De manera similar, otro transductor convertirá la señal de salida a la forma apropiada en el lugar de destino. Un micrófono y un altoparlante son ejemplos típicos de transductores de entrada y de salida, respectivamente, en el caso de mensajes analógicos. Por otra parte, el teclado de una máquina telex y una impresora son ejemplos de transductores de mensajes digitales.

Resulta claro que también hablaremos de señales analógicas y digitales. Las primeras serán magnitudes eléctricas que variarán en forma continua en función del tiempo. Las segundas, en cambio, tomarán un número finito de valores.

En este libro estudiaremos fundamentalmente la comunicación mediante señales eléctricas digitales en las redes de telecomunicaciones, es decir, comunicaciones digitales. En forma particular analizaremos la transmisión de señales eléctricas digitales por medio de los sistemas eléctricos de comunicación.

Aunque algunas señales son digitales en naturaleza (teletipo, salida de computadora, etc.), muchas de ellas son analógicas y si las queremos transmitir digitalmente deben muestrearse previamente con cierta tasa periódica y convertirse en muestras discretas de amplitud mediante un proceso de cuantificación, para que luego puedan codificarse para su transmisión eficiente.

Voz, video, facsímil, etc., son ejemplos de señales analógicas transmitidas digitalmente previo proceso de muestreo, cuantificación y codificación. Este proceso de conversión analógico-digital lo estudiaremos más adelante.

Los elementos de un sistema completo de comunicación se muestran en la figura I.1. Excluyendo los transductores hay tres partes esenciales: el transmisor, el canal o vínculo de transmisión y el receptor.

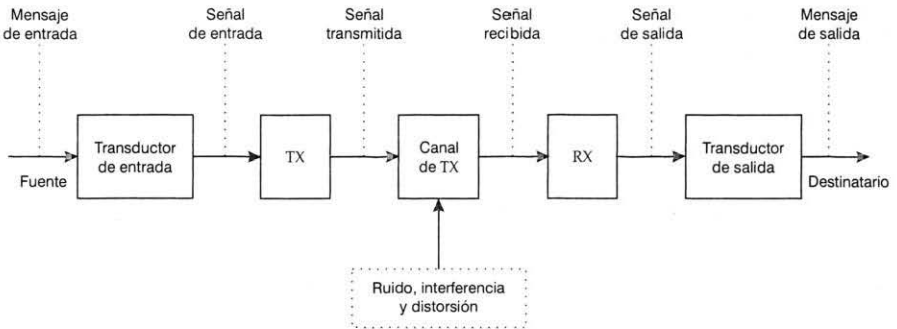


Figura I.1. Elementos principales de un sistema de comunicación

El transmisor acopla el mensaje en el canal en forma de una señal transmitida y, a efectos de una transmisión eficiente, varias operaciones de procesamiento de la señal pueden ser realizadas. La más común es la modulación donde se adapta la señal transmitida a las propiedades del canal por medio de una onda portadora.

El canal de transmisión o medio es la conexión eléctrica entre el transmisor y el receptor, que puede ser un par de conductores, un cable coaxial, un radioenlace, una fibra óptica, etcétera.

El receptor extrae la señal que se desea del canal y la entrega al transductor. La operación clave del receptor es la demodulación, o sea el proceso inverso de modulación del transmisor, que restaura la señal a su forma original.

En el trayecto de transmisión de la señal ocurren ciertos efectos indeseables, que aparecen como alteraciones de la forma de la señal. Estas contaminaciones que se introducen en todo el sistema, por lo general se las refiere al canal considerando el transmisor y el receptor como ideales. Estas contaminaciones son: la distorsión, la interferencia y el ruido.

En capítulos posteriores analizaremos en detalle el efecto de estas contaminaciones sobre la exactitud con que se recibe la señal digital. Es oportuno, sin embargo, que definamos cada una de ellas.

La distorsión es una alteración en la forma de la señal debido a la respuesta imperfecta del sistema. Sus características distintivas son que sólo se produce en presencia de la señal y que puede reducirse tanto como se desee, aunque muchas veces esto resulta antieconómico.

La interferencia es una contaminación que se produce por la aparición de señales externas al sistema, generalmente producidas por el hombre. Si bien en teoría la solución a la interferencia es simple, es decir, la eliminación de la fuente perturbadora, en la práctica no siempre es factible.

El ruido está constituido por señales eléctricas aleatorias impredecibles debidas a causas de origen natural, tanto externas como internas. A diferencia de las dos contaminaciones anteriores, no puede eliminarse ni en teoría, por lo que constituyen una de las limitaciones físicas fundamentales de las comunicaciones eléctricas. En este punto debemos aclarar que las contaminaciones mencionadas resultan más nocivas cuando la señal recibida está más debilitada por efecto de la atenuación del medio de transmisión. La atenuación es una característica de todos los medios de transmisión y representa el debilitamiento progresivo de la potencia de la señal con la distancia. En consecuencia, más que hablar de valores de ruido en forma absoluta es preferible hablar en términos de relación señal a ruido, simbolizado por S/N .

Hemos mencionado el ruido como una de las limitaciones físicas fundamentales; la otra es el ancho de banda. Ambas determinan categóricamente qué proyectos son factibles aun antes de plantearnos la implementación tecnológica. La noción de ancho de banda se aplica tanto para las señales como para los sistemas.

En el caso de las señales nos indica su velocidad de variación y, en consecuencia, la cantidad de información que contiene. Así, una señal vocal tendrá un ancho de banda de unos 4 KHz, una señal de música de alta fidelidad superará los 20 KHz y una de video excederá los 4 MHz.

Por otro lado, la aptitud de un sistema para seguir las variaciones de la señal inyectada a su entrada también se expresa como ancho de banda de transmisión. En efecto, debido a los elementos almacenadores de energía, los sistemas de comunicaciones presentan cierta "inercia" que dificulta la transmisión de señales que varían con rapidez. Por lo tanto, para evitar excesivas distorsiones durante la transmisión, es necesario que el espectro de la señal sea inferior al ancho de banda del sistema.

Ambas limitaciones, ruido y ancho de banda, determinan la capacidad del sistema que, según Shannon (capítulo II), no debe ser excedida por la tasa de información. La capacidad del sistema se expresa en la conocida ley de Shannon-Hartley

$$C = B \log(1 + S/N)$$

donde B es el ancho de banda del sistema. Esta ley permite establecer el límite superior del comportamiento del sistema.

El objetivo de un sistema de comunicación digital es transmitir mensajes lo más rápidamente posible y con una cantidad mínima de errores. Puesto que la capacidad de los sistemas es finita, debe encontrarse un compromiso entre el tiempo de transmisión, potencia transmitida, ancho de banda y relación señal a ruido; compromiso que, además, no puede soslayar las limitaciones de la tecnología.

Para acercarnos al objetivo establecido contaremos con dos operaciones: la modulación y la codificación. Varios capítulos de este libro están dedicados a estudiarlos. Por ahora cabe decir que el éxito de un proyecto dependerá, entre otras cosas, de la correcta elección del tipo de modulación y de la codificación adecuada. Tal como se establece en el llamado teorema fundamental de la teoría de la información [1]:

Si la tasa de información de una fuente no excede la capacidad de un canal de comunicación, entonces existe una técnica de codificación tal que la información puede transmitirse por el canal con una probabilidad de error arbitrariamente pequeña a pesar de la presencia del ruido.

1.2. UN POCO DE HISTORIA

Las invenciones, en su gran mayoría, surgen para satisfacer necesidades, para encontrar solución a problemas existentes. Sin embargo, para concretar un invento no basta la existencia de la necesidad sino que deben concurrir conjuntamente otros dos hechos esenciales. En primer lugar debe contarse con la base de conocimientos teóricos que sustente la invención y, además, debe poseerse y manejarse la tecnología capaz de implementarla.

En el área de las telecomunicaciones encontramos un buen número de realizaciones que corroboran esta afirmación, así como numerosos proyectos relegados durante años hasta que se dan estos tres factores.

Sería demasiado simplista reducir la historia del desarrollo de las comunicaciones a la mera reseña de valiosos descubrimientos aportados por eminentes científicos e investigadores. Aspectos como las circunstancias políticas, las relaciones económicas, los conflictos bélicos, las situaciones sociales, etc., son ejemplos insoslayables cuando se desea considerar lo que ha llevado a esos hombres a abordar un problema en particular.

Paradójicamente, muchas invenciones han llevado, a su vez, a formular nuevos e insospechados problemas. El comportamiento de determinados desarrollos no era, algunas veces, el que se esperaba teóricamente. Este hecho era, en primera instancia, atribuido a fallas en la implementación, a imperfecciones que podrían subsanarse optimizando el proyecto. En consecuencia, llevaba mucho tiempo reconocer que había una carencia teórica y, por lo tanto, remediarla.

Situaciones como éstas se han presentado reiteradamente a lo largo de la historia, por lo que no debemos pensar que no se presenten hoy en día. Es muy probable que muchas de las limitaciones que se experimentan en nuestros sistemas actuales sean manifestaciones de fenómenos que no se tienen en cuenta o no están correctamente fundamentados.

Paradójicamente, los primeros y rudimentarios sistemas de comunicación fueron digitales, o sea que lo que hoy vemos como una innovación que ha modificado drásticamente la configuración de las redes, el tipo de servicios y el diseño de los equipos no es más que un retorno; como suele suceder, detrás de cada paradoja suele haber una explicación. Con una tecnología rudimentaria y con escasas bases teóricas, el hombre de la antigüedad no tuvo más remedio que intentar transmitir un número discreto de eventos para facilitar en el lugar de destino un reconocimiento (hoy diríamos una detección) más preciso. Así pareció entenderlo Polibius, el general griego que en el año 300 a. C. diseñó un sistema basado en un arreglo de diez antorchas en cada estación repetidora en la que se construyeron dos paredes, separadas por más de un metro, de dos metros de largo y de altura conveniente. Sobre cada una de las paredes se ubicaban cinco antorchas y las letras se transmitían mediante el siguiente código:

Tabla I.1. Código usado por Polibius (antorchas encendidas)

		<i>Pared izquierda</i>				
		1	2	3	4	5
<i>Pared derecha</i>	1	A	B	Γ	Δ	E
	2	Z	H	Θ	I	K
	3	Λ	M	N	Ξ	O
	4	Π	Π	Σ	T	Ψ
	5	Φ	Ξ	Ψ	Ω	

Así, por ejemplo, la letra alfa estaba representada por una antorcha en cada pared, la beta por una antorcha en la pared de la derecha y dos en la izquierda, etcétera.

Los primeros intentos de comunicaciones eléctricas se efectuaron en el siglo XVIII. Fueron también digitales y se basaban en tensiones altas y bajas corrientes. Así, el telégrafo de C. M. (sólo se conocen las iniciales) propuesto en el *Scots' Magazine* del 17 de febrero de 1753 utilizaba un generador electrostático que actuaba como transmisor, y se conectaba, según se deseaba, a 26 líneas de transmisión paralelas, cada una de las cuales representaba una letra. El receptor consistía en bolitas de médula suspendidas frente a cada línea por medio de un hilo de seda. La bolita que resultaba repelida identificaba la letra transmitida. Como puede observarse, la codificación era elemental, lo cual obligaba a un sistema de transmisión complejo. De todos modos, la falta de una necesidad perentoria relegó el desarrollo del sistema y no existen pruebas fehacientes de que se haya llevado a la práctica.

A fines del siglo XVIII la efervescencia de la Revolución Francesa demandará sistemas de comunicaciones más veloces. El francés Claude Chappe intentó desarrollar un sistema de telégrafo electrostático, pero la falta de un soporte tecnológico adecuado lo hizo optar por el antiguo método del general griego, aunque ahora mejorado por los aportes de la física y la mecánica. Así nació el semáforo visual, que iba a expandirse rápidamente por todo el imperio napoleónico. El uso estratégico que hizo Napoleón de este sistema es lo que, precisamente, dio un impulso a la investigación sobre el telégrafo eléctrico. En la primavera de 1809 Napoleón reconquistó Munich en sólo seis días, gracias a que con el semáforo de Chappe se le avisó que los austriacos habían cruzado el Inn e invadido Baviera, forzando al rey Maximiliano a abandonar la ciudad para refugiarse en Dilligen. Admirado por la efectividad del sistema, el monarca, ya restituido en Munich, solicitó a la Academia Bávara de Ciencias que presentara propuestas para instalar el telégrafo en el reino.

De esta manera, S. T. Sömmering presentó en 1811 un telégrafo eléctrico basado en el C. M., pero reemplazando el generador electrostático por una pila voltaica (inventada en 1800) y usando como receptor un dispositivo electroquímico. Precisamente este tipo de detector le restó velocidad al sistema y diez años más tarde, gracias a que H. C. Oersted descubrió el electromagnetismo, le fue posible a Ampere proponer el uso de un detector electromagnético.

Con la aparición del ferrocarril en 1830 la necesidad de las telecomunicaciones aumentó. Mientras tanto, Ohm había establecido en 1827 la relación entre tensión, resistencia y corriente, con lo cual comenzó a ser evidente que el límite de distancia de un sistema debía estar dado por la sensibilidad del receptor, la resistencia de la línea y la máxima diferencia de potencial aplicada.

En 1837 Wheatstone y Cooke instalaron un sistema de demostración en un tramo de la línea férrea Londres-Birmingham: una versión mejorada del telégrafo de Schilling, que utilizaba seis líneas de transmisión y un detector con cinco agujas magnéticas de las cuales dos por vez señalaban la letra transmitida. Obsérvese que se trata de un sistema de codificación que permite, mediante transmisión paralelo-simultánea, la reducción del número de líneas de transmisión.

Un año más tarde, en 1838, en el ferrocarril Nuremberg-Furth, en Alemania, se instaló un sistema telegráfico basado en el desarrollo de Gauss-Webber, que requería un solo par de conductores mediante el uso de un código secuencial de señales positivas y negativas. Debe notarse que los dos tipos de código, serie y paralelo, se encuentran en uso hasta el presente. El código de Wheatstone-Cooke es lo que hoy llamamos código de relación constante, es decir un código en el cual el contenido relativo de marcas y espacios es fijo (2 a 3 en este caso). Con el objeto de reducir el número de líneas de transmisión se desarrollaron códigos secuenciales, y para aumentar la velocidad se asociaron las palabras de código más cortas con las letras más frecuentes. Éstos fueron los primeros códigos de longitud variable que culminaron con el conocido código de Morse (1843) que tal vez merecería con más justicia denominarse código de Vail dado el talento e ingenio desplegado por Alfred Vail para hacer operable el complicado sistema diseñado por Samuel Morse.

Para solucionar la limitación impuesta por la línea de transmisión se introdujeron dispositivos regeneradores en puntos intermedios; eran activados por la señal antes que se degradase irremediablemente y permitían producir una nueva señal. Fueron varios los sistemas desarrollados basados en esta idea que recibieron el descriptivo nombre de *relay*; baste citar los de Cook-Wheastone (1837), Edward Davy (1838), Morse (1840), etcétera.

La primera demostración comercial del telégrafo de Morse se llevó a cabo en 1844, sobre una línea de 40 millas entre Washington y Baltimore.

Con la difusión del sistema telegráfico comenzó a intentarse el tendido de los primeros cables submarinos: Connecticut (1849), Dover-Calais (1850), y entre Inglaterra y Estados Unidos en 1858. La sensibilidad de los detectores aumentó notablemente, lo cual permitió incrementar la velocidad de transmisión. No obstante, las altas capacidades y las bajas impedancias características de los cables producían comportamientos diferentes al de las líneas aéreas. Las velocidades de transmisión no lograban superar las 20 palabras por minuto.

En 1855 Kelvin presenta lo que sería el punto de partida en el estudio teórico de las líneas de transmisión, la denominada Ley KR (K, por capacidad). En 1887 Olivier Heaviside completa el estudio introduciendo las constantes G y L.

Con la invención del teléfono por A. G. Bell (1876), a partir de la década de 1880 el desarrollo de la telefonía domina el campo de las telecomunicaciones. Las redes se tornan básicamente analógicas y la mayoría de los servicios digitales debieron adaptarse a esta forma de transmisión. La introducción del teletipo (1931) no modificó esta situación y en la mitad de nuestro siglo el mundo de las telecomunicaciones era esencialmente analógico, salvo los circuitos telegráficos transoceánicos y algunos sistemas de radio. Durante ese periodo los trabajos teóricos de H. Nyquist, R. Hartley, J. Johnson y C. Shannon, entre otros, habían aclarado por completo los conceptos de transmisión, ruido, información, etc., sentando las bases teóricas para el gran salto: el de las modernas comunicaciones digitales.

Las modernas comunicaciones digitales engloban dos corrientes confluentes: la de las señales analógicas digitalizadas y la de transmisión de datos. La primera nos remonta a 1862 cuando Reis intentó construir un teléfono basado en un contacto de platino que abría o cerraba según el movimiento de un diafragma. En los albores del siglo XX Miner (1903) logró desarrollar un modulador PAM muestreando una señal telefónica a 4 KHz. Sin embargo la real codificación digital de la señal analógica la realizó A. Reeves en 1936-1937, quien aparte de cuantificar la señal muestreada codificó las muestras en un código binario, lo cual condujo al desarrollo del PCM.

En 1947 la Bell Telephone Laboratories reportó la operación del primer sistema PCM. En el mismo tiempo Deloraine en Francia inventa la modulación delta. A pesar de estar desarrollado, el sistema PCM esperó durante casi 15 años para que la necesidad de su utilización fuera perentoria y hubiera una tecnología apta para su implementación.

La invención del transistor (1948) y su posterior desarrollo y aplicación como dispositivo de conmutación proveyó la tecnología requerida. El crecimiento de la red telefónica en las grandes urbes y las dificultades en el tendido de nuevos cables creó la necesidad de utilizar las líneas interficina existentes en forma más intensa a través del multiplexado de tiempo.

La revolución de la computación tuvo un doble efecto en el crecimiento de las comunicaciones digitales: en primer lugar, la necesidad de transmitir datos creció vertiginosamente, desarrollándose en 1958 los primeros sistemas de transmisión a larga distancia para fines militares, e irrumpiendo la comunicación a altas velocidades en los comienzos de la década de 1960. En segundo lugar, la tecnología de la computación con sus compuertas digitales integradas y sus memorias pasó a ser aplicada en los desarrollos de las telecomunicaciones, proveyendo equipos a costos cada vez más competitivos con sus similares analógicos. La producción comercial de circuitos integrados comenzó en 1961, y en 1964 se puso en servicio la primera central de conmutación totalmente electrónica

(No. 1 ESS). Los desarrollos sobre fibra óptica, LSI y trabajos teóricos sobre ecualización adaptiva y codificación para el control de errores son historia reciente.

Las perspectivas parecen indicar que la historia de las telecomunicaciones, que comenzaron con experiencias digitales para luego desarrollarse analógicamente, se encamina hacia una red totalmente digital que reconoce los estados discretos como los más aptos para vencer las limitaciones físicas.

Si bien se requirieron 20 siglos para ir de la comunicación de datos por antorchas a la transmisión de señales eléctricas, bastaron sólo 20 años para evolucionar de la transmisión de datos primitiva a la avanzada comunicación y procesamiento a alta velocidad, y la historia aún continúa...

I.3. VENTAJAS DE LAS REDES DIGITALES

Como hemos mencionado, las redes de telecomunicaciones fueron evolucionando hasta los años cincuenta primordialmente para permitir la comunicación analógica de la voz humana. Desde entonces y gracias a los desarrollos de la microelectrónica ha sido factible reemplazar técnica y económicamente esta red analógica por modernos sistemas digitales. La comunicación digital permite, en esencia, que cualquier tipo de información —voz, datos, imágenes— sea convertida en pulsos digitales y transmitida por medio de éstos.

A continuación describiremos las diversas ventajas que ofrecen las redes digitales comparadas con las analógicas. El orden en que se presentan no implica un rango de importancia rígido ya que su relevancia depende de criterios subjetivos y del tipo de servicio que prestan. Así, por ejemplo, la capacidad criptográfica de los sistemas digitales es de relativa importancia para los circuitos comerciales pero fundamental en aplicaciones militares.

Es oportuno señalar que algunas de las ventajas que enunciaremos son de aplicación práctica sólo para una red totalmente digital. El caso citado de codificación criptográfica tiene sentido solamente si la hermeticidad del mensaje se establece en la fuente de origen y recién se descifra en el destinatario. Es decir que se requiere un medio de transmisión transparente, lo que implica un sistema que opere de extremo a extremo sin considerar la naturaleza del tráfico que transmite.

Comencemos por enunciar las diversas ventajas de las redes digitales:

- Facilidad para diversos servicios
- Facilidad de multiplexación
- Facilidad para la utilización de tecnología moderna
- Integración de la transmisión y la conmutación
- Regeneración de la señal
- Funcionamiento con bajas relaciones señal-ruido
- Facilidad de señalización
- Facilidad criptológica
- Facilidad de monitoreo

I.3.1. Facilidad para diversos servicios

La posibilidad de utilizar la red digital para la comunicación de cualquier forma de información es, a nuestro juicio, su ventaja potencial más trascendente; la evolución de las redes telefónicas a una red digital de servicios integrados —RDSI— (o ISDN por Integrated Service Digital Network) permitirá la integración de diversos servicios.

Así, podemos reflexionar que la información digital cada día desempeña un papel más preponderante en el quehacer cotidiano, ya sea desde el campo económico o científico hasta el ámbito personal y familiar; cualquier señal de datos que se desee transmitir encuentra cierto número de dificultades. Por ejemplo, el usuario de una computadora hogareña se encuentra limitado a transmitir o recibir a la velocidad que le fija su línea telefónica, cuyo ancho de banda de 4 KHz está diseñado en función de las señales vocales, por lo que está imposibilitado a tener acceso a servicios que requieran mayores tasas de información, como señales de video.

En otras palabras, el hecho de que la única red global de comunicaciones que existe, la red telefónica internacional, haya sido diseñada principalmente para la transmisión de señales vocales y no para señales digitales de datos, inhibe a la mayoría de los usuarios para acceder a diversos servicios que no sean telefónicos y obliga a éstos a adaptar sus señales para asemejarlas a las señales vocales.

Un principio de solución son las denominadas redes de área local (LAN por Local Area Network), en las que mediante un vínculo se interconectan computadoras, terminales y otros periféricos dentro de un edificio o grupo de edificios. Dicha red admite diversas configuraciones y diversas capacidades. El vínculo puede ser desde el simple cable plano hasta un coaxial o incluso una fibra óptica. Obviamente el alcance de esta red es limitado y restringido, y sólo satisface plenamente una necesidad muy específica.

El objetivo de las empresas de telecomunicaciones es el desarrollo de una red de esas características pero con cobertura universal; es decir, una red que permita a cualquier usuario comunicarse mediante el servicio que desee con cualquier otro usuario en cualquier instante y en cualquier lugar. Para ello el usuario sólo requerirá un nuevo acceso a la red a través de la cual podrá acceder tanto a servicios digitales como analógicos, tales como: bases de datos, conexiones de computadora, teletex, facsimil, televisión, teleconferencia, videoconferencia, telefonía (obviamente) y otros.

El rango de aplicación de las RDSI se extenderá desde las bajas velocidades, 300 bits/seg¹ o menos, hasta velocidades que superen los 100 Mbits/seg. Las más bajas son para los servicios de telemedición, seguridad, etc.; velocidades un poco mayores permiten el manejo interactivo de datos. Las señales vocales se codifican usualmente a una tasa de 64 Kbits/seg y, como luego veremos, esta elección ahora tiene una influencia determinante en los formatos y jerarquías del resto de los servicios.

También a 64 Kbits/seg suelen proveerse aplicaciones de imágenes fijas tales como facsímil, gráficos o televisión cuadro a cuadro.

La transmisión de archivos de datos o la interconexión entre computadoras con procesamiento distribuido requiere mayores velocidades (1.5 Mbits/seg o más). La transmisión de video, tanto para entretenimiento como para teleconferencia, requiere las velocidades más altas. La codificación directa de una señal convencional de alta calidad demanda alrededor de 100 Mbits/seg, valor que puede reducirse con técnicas de codificación adecuadas que aprovechan la redundancia de la señal.

Un beneficio adicional de la red digital integrada es que el tráfico que proviene de distintas fuentes puede “mezclarse” en un único medio de transmisión. En los sistemas analógicos esto es muy difícil de lograr, no sólo por los diversos requerimientos de ancho de banda sino por los distintos niveles de calidad; por ejemplo, las señales de TV requieren mayor calidad que las señales vocales, lo que debe tenerse muy en cuenta cuando se combinan televisión y canales vocales FDM, en sistemas de transmisión analógica de banda ancha.

¹ Para la definición de bits/seg véase el capítulo II.

1.3.2. Facilidad de multiplexación

Como se sabe, el proceso de multiplexación permite compartir un mismo medio de transmisión entre varias señales. De esta manera se realiza un mejor aprovechamiento del vínculo y puede decirse, sin lugar a dudas, que sin esta técnica el vertiginoso desarrollo de las comunicaciones no hubiera sido posible. Como contrapartida los equipos terminales se encarecen, lo que constituye un compromiso que permanentemente enfrentan los ingenieros en telecomunicaciones, el costo del medio de transmisión *versus* el costo de los equipos multiplexores de las estaciones terminales.

Los primeros sistemas de multiplexación fueron por división de frecuencia (FDM), pero día a día los multiplexores por división de tiempo (TDM) resultan más y más económicos. Si la señal a multiplexar ya se encuentra digitalizada, esta comparación de costos resulta aún más favorable para los TDM. También se acentúa esta tendencia a medida que subimos en la jerarquía de los niveles de multiplexación, es decir, cuando agrupamos más señales.

En general, cuando se habla de jerarquías nos estamos refiriendo a señales vocales, por lo que se habla de sistemas de 24, 30, 120 o más canales telefónicos. Sin embargo, la facilidad de intercalar en el flujo de bits que representa a esos canales telefónicos señales provenientes de otros tipos de servicios constituye, como hemos mencionado en el párrafo anterior, el hecho fundamental que conduce a la integración de servicios.

Cabe consignar que FDM y TDM no son las únicas formas posibles de multiplexar señales, hay otras técnicas basadas en la asimetría de los portadores. Sin embargo, solamente la multiplexación en frecuencia y tiempo ha encontrado, hasta ahora, aplicación práctica y generalizada.

1.3.3. Facilidades para la utilización de tecnología moderna

La aplicación de la tecnología electrónica moderna ofrece grandes ventajas a los sistemas de comunicaciones en general, pero especialmente a los sistemas digitales. Por ejemplo, un multiplexor o una matriz de conmutación para señales digitales en división de tiempo se realiza con los mismos circuitos básicos que se utilizan en la construcción de computadoras digitales, es decir, compuertas lógicas y memorias. Concretamente, el punto de cruce básico de una matriz de conmutación digital no es nada más que una compuerta AND a la cual se ha asignado una entrada lógica para la señal de mensaje y las otras para el control, es decir, para la selección del punto de cruce. Por lo tanto, resulta claro que los desarrollos de la tecnología de los circuitos integrados digitales para circuitos lógicos de computación y memoria se aplican directamente en los equipos de conmutación y transmisión digital. Pero si bien muchos circuitos integrados desarrollados para usarse en computadoras se aplican directamente en telecomunicaciones, también se han desarrollado circuitos integrados en gran escala (LSI) para el uso específico en equipos de comunicaciones. Ambos se aplican a las principales funciones de los sistemas, con lo que el tamaño y el costo de los equipos se reduce notablemente, y aumenta además la confiabilidad de la red. La comparación de costos resulta aún ventajosa, pero incluye el inconveniente de la conversión analógico-digital y viceversa, por ejemplo, cuando las matrices de conmutación digitales operan en un medio de transmisión analógico. Obviamente que cuando se asocian sistemas de conmutación digital con enlaces digitales se obtienen reducciones significativas de complejidad y costo. El bajo costo relativo y el buen desempeño de los circuitos digitales permiten aplicaciones que serían muy onerosas con componentes analógicos; un ejemplo lo constituyen las matrices de conmutación sin bloqueo, que de hecho no pueden realizarse analógicamente, salvo para tamaños pequeños.

Los beneficios de la tecnología de dispositivos modernos no sólo son privativos de los circuitos digitales. Los circuitos integrados analógicos también han progresado significativamente, de mane-

ra que las implementaciones analógicas tradicionales han mejorado en forma notable; sin embargo, uno de los requisitos primordiales de los componentes analógicos es la linealidad, por lo que la manufactura de los circuitos digitales resulta más sencilla.

Las implementaciones digitales tienen una ventaja funcional inherente sobre las analógicas. Esta ventaja proviene de la relativa facilidad con que las señales digitales pueden ser multiplexadas. La mayor limitación con el uso de los componentes LSI viene de la limitada disponibilidad de conexiones externas. Con técnicas TDM un solo contacto físico puede usarse para el acceso de un canal múltiple. Así, la misma técnica ya usada para reducir costos en sistemas de transmisión puede usarse dentro de un módulo local para minimizar las interconexiones y maximizar la utilización de integración en gran escala. De esta manera se llega a obtener una matriz de conmutación en un *chip* mediante la multiplexación de un gran número de canales sobre un pequeño número de conexiones externas.

Dos tecnologías de transmisión muy utilizadas en la actualidad son los satélites y la fibra óptica. Aunque muchos sistemas de satélite usan multiplexores por división de frecuencia analógicos, las ventajas del modo de acceso múltiple por división de tiempo digital (TDMA) señalan que en el futuro los satélites serán digitales.

La interfaz electrónica para la fibra óptica (lasers, diodos emisores de luz, fotodiodos) tiene básicamente un modo de operación *on-off* (no lineal). Aunque la modulación por amplitud o ancho del pulso puede usarse por estos dispositivos para dar una transmisión analógica, el énfasis del desarrollo de las fibras ópticas es claramente digital. Así, en el caso de fibra óptica el enlace de transmisión implica un modo digital de operación.

Otra área de aplicación de la tecnología digital es el procesamiento digital de la señal. Se trata de una operación sobre una señal para realzar o transformar sus características. El procesamiento digital de una señal puede aplicarse a formas de onda digitales o analógicas, amplificación, ecualización, modulación y filtrado son ejemplos comunes de procesamiento de señales.

Por procesamiento de señales digitales entendemos el uso de la lógica circuital y circuitos aritméticos para implementar funciones de procesamiento. Algunas veces las señales analógicas se digitalizan con el único propósito de procesarlas digitalmente, luego de lo cual se convierten nuevamente en analógicas.

Las principales ventajas del procesamiento digital de señales son:

- La inmunidad de los circuitos digitales a las pequeñas imperfecciones y a los elementos parásitos, por lo que pueden producirse con características estables, sin ajustes finos y sin tolerancias de envejecimiento.
- Una estructura básica única puede usarse para diversos tipos de señales y aplicaciones con sólo modificar un algoritmo o un parámetro en una memoria virtual.
- Un único circuito de procesamiento puede utilizarse para diversas señales mediante el almacenamiento de los resultados de cada proceso en una memoria de acceso aleatorio.
- Puesto que las entradas y salidas de los circuitos de procesamiento son señales digitales, pueden desarrollarse periódicamente evaluaciones automáticas de funcionamiento mediante la comparación con estructuras de datos predeterminadas almacenadas en la memoria.
- Por la facilidad de los circuitos digitales para tomar decisiones lógicas, los procesadores de señales digitales pueden efectuar diversas funciones de imposible o impracticable realización en implementaciones analógicas.

I.3.4. Integración de la transmisión y la conmutación

Los equipamientos de transmisión y conmutación de las redes telefónicas han sido por tradición sistemas prácticamente independientes. Si bien era necesario e imprescindible especificar con precisión la interfase entre ambos, sus límites se establecían justo en los repartidores de las centrales.

Con el desarrollo de los sistemas TDM y los equipos de conmutación digital, cada vez se fue haciendo más evidente la similitud entre ambos y con ello aumentó la posibilidad y conveniencia de integrarlos.

En realidad, y como se verá en los capítulos posteriores, la primera etapa de los sistemas de conmutación digital genera el primer nivel del multiplexor por división de tiempo. En consecuencia, ambas etapas pueden integrarse fácilmente en una única operación. En la figura I.2 se esquematizan ambas situaciones, resultando evidente la reducción de equipamiento y la ausencia de un límite definido entre el equipamiento de transmisión y el de conmutación.

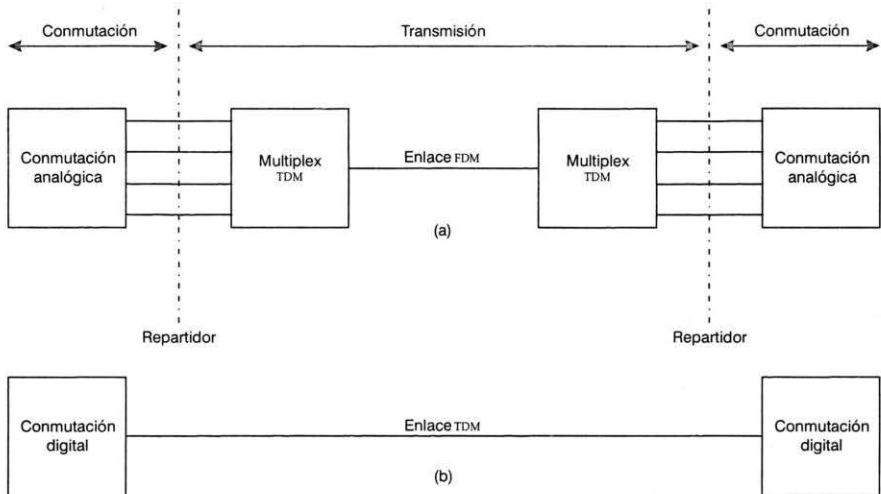


Figura I.2. a) Transmisión y conmutación no integradas, b) transmisión y conmutación integradas

Debe notarse, además, que las señales vocales se digitalizan al acceder a la central de conmutación digital y permanecerán digitalizadas hasta llegar a su destino, a menos que tengan que transitar por algún sistema analógico. En consecuencia, la integración de la transmisión y la conmutación no sólo implica un ahorro importante de equipamiento sino que mejora la calidad de la señal recibida en la medida en que tendremos para el caso óptimo de digitalización total una única conversión analógica-digital y una única conversión digital analógica.

De esta forma la calidad del enlace queda prácticamente determinada por el proceso de cuantificación si es que la línea de transmisión no introduce un número significativo de errores.

Otra ventaja que tenemos que consignar es que en una red totalmente digitalizada la calidad de los circuitos de larga distancia no diferiría de los circuitos locales presentando iguales características de nivel de señal, distorsión y ruido.

1.3.5. Regeneración de señal

Hemos dicho que un mensaje digital es una secuencia ordenada de símbolos seleccionados de un conjunto finito de elementos discretos. Precisamente esa limitación del número de elementos posibles que puede esperar un receptor es lo que le permite operar con mayor facilidad que los sistemas analógicos.

A los receptores, como a los seres humanos, les resulta más sencillo reconocer estados perfectamente definidos. En la medida en que las contaminaciones que pretenden alterar esos estados son de naturaleza analógica, basta fijar límites precisos dentro de los cuales se reconozca cada estado para que el receptor pueda discernir entre ellos.

Obviamente esa toma de decisión deberá hacerla el receptor antes de que la contaminación sea relativamente tan grande que produzca el desplazamiento de un estado fuera de sus límites de existencia; en ese caso el receptor estará cometiendo un error.

Una manera efectiva de evitar esos errores consiste en colocar receptores intermedios que reconozcan los estados antes que la señal se degrade en demasía. Una vez realizado ese reconocimiento, podrá transmitirse el mensaje nuevamente, libre ya de contaminaciones y con un nivel adecuado para que llegue hasta otro receptor. Estos receptores intermedios en conjunto con sus transmisores asociados reciben el nombre de repetidores regenerativos o simplemente regeneradores.

Meditese con atención que la regeneración sólo es posible si trabajamos con estados discretos, es decir, con una colección de símbolos finita que permite fijar intervalos de existencia a cada símbolo. Ello no es posible en el caso analógico, donde todos los valores son posibles y, por lo tanto, cualquier degradación es irreversible.

Un atributo fundamental de los sistemas digitales es que la probabilidad de errores en la transmisión puede hacerse arbitrariamente pequeña insertando repetidores regenerativos en puntos intermedios del enlace. Si la distancia entre ellos es adecuadamente próxima, los regeneradores pueden recrear la señal digital original antes de que el ruido, la interferencia y la distorsión sean tan elevados que produzcan errores de decisión.

De esta manera la señal regenerada resulta libre de degradaciones, a diferencia del caso analógico donde las contaminaciones se acumulan a lo largo del enlace. Este hecho se produce en los enlaces analógicos aun cuando haya repetidores que no sean más que simples amplificadores que no sólo elevan el nivel de la señal sino el de las contaminaciones que se le han adicionado. Esto obliga a un cuidadoso diseño y control de los enlaces analógicos, teniendo en cuenta la calidad de extremo a extremo de la red de la que pueden llegar a formar parte.

Como contrapartida, un enlace digital se diseña solamente a partir de especificar una probabilidad de error determinada, lo cual es normalmente factible si se provee un adecuado número de regeneradores.

En estas condiciones, y cuando el enlace digital se utiliza para transmitir señales analógicas que han sido digitalizadas, la calidad de la señal recibida queda determinada primordialmente por el proceso de conversión análogo-digital.

1.3.6. Funcionamiento con bajas relaciones señal-ruido

En el párrafo anterior describimos cualitativamente la relativa facilidad con que el receptor puede discriminar estados discretos. Como veremos al tratar la transmisión en banda base, basta una relación señal a ruido del orden de los 20 dB para obtener menos de un error por cada millón de dígitos transmitidos. Este hecho no es en realidad tan promisorio como parece a simple vista, pues como veremos más adelante, las exigencias de ancho de banda de los sistemas digitales son muy

grandes, por lo que el ruido del sistema también es alto. Dicho de otra manera, la facilidad de los sistemas digitales para operar a bajas relaciones señal-ruido es más un requerimiento que una ventaja.

Si tomamos el caso concreto de señales vocales digitalizadas, veremos que el requerimiento de ancho de banda es ocho veces mayor que para la señal analógica. Si el medio de transmisión es, por ejemplo, un par de cables, tendremos mayor atenuación de la señal, por lo que el sistema digital debe ser capaz de operar a más bajas relaciones señal a ruido para poder ser una alternativa viable.

Ya que nos referimos al caso de telefonía, cabe mencionar algunas ventajas que los sistemas digitales presentan desde el punto de vista subjetivo. En efecto, tanto el ruido como la interferencia son particularmente molestos durante las pausas de las conversaciones telefónicas. Niveles de ruido que permanecen imperceptibles durante las alocuciones producen notables molestias durante los silencios. Esta consideración, junto con pruebas subjetivas, lleva a especificar relaciones señal-ruido del orden de los 45 dB.

En los sistemas digitales, en cambio, las pausas son codificadas de igual manera que en cualquier otro nivel de señal vocal. En consecuencia se transmiten con presencia de señal, y se les regenera eliminando todo el ruido introducido en el medio de transmisión, con ello la condición de reposo queda fijada por el proceso de codificación y no por el medio de transmisión.

Idénticas consideraciones pueden hacerse respecto al rechazo de la interferencia y la diafonía, factores que suelen ser más molestos que el mismo ruido aleatorio. La diafonía, es decir, la perturbación entre canales vecinos, es más molesta aun cuando es inteligible y cuando se produce en las pausas de la conversación telefónica. Nuevamente en los sistemas digitales no solamente se elimina en el proceso de regeneración sino que, en ocasiones, sólo puede producir algún error que aparece aleatoriamente y, en consecuencia, resulta ininteligible.

La capacidad de los sistemas digitales de ser bastante inmunes a la interferencia y a la diafonía los hace particularmente aplicables en áreas congestionadas. Por eso, la conveniencia de los sistemas digitales es innegable en zonas donde se reutilizan las mismas frecuencias intensivamente para sistemas radioeléctricos, o donde se indica la utilización de polarización cruzada.

1.3.7. Facilidad de señalización

La facilidad que los sistemas de transmisión digital presentan para incorporar información de control asociada a los circuitos de voz, fundamentalmente se debe a que dicha información es sobre todo digital. En efecto, tanto la indicación de estados (colgado, descolgado) como la información de los dígitos discados y la tarificación son digitales.

Una forma de incorporar la información de control en un enlace digital es asignándole un canal común donde se multiplexan las señalizaciones de los diversos canales, otra forma consiste en asignar bits de control junto con el mensaje en cada canal individual. Pero independientemente de los métodos usados, la información de control una vez intercalada se confunde como unos bits más en el flujo total de datos, quedando obviamente a cargo del receptor la tarea de reconocerlos y decodificarlos.

En contraste, la situación no es tan simple en los sistemas analógicos donde existen diversos procedimientos para ingresar las señales de control. Así se encuentran señalizaciones dentro y fuera de banda, con un solo tono o multitono, a alto nivel o bajo nivel, etc. Esta diversidad de formatos lógicamente implica serias complicaciones de compatibilidad y en muchos casos la necesidad de proveer interfaces adecuadas.

En conclusión, los sistemas digitales permiten que la información de control se inserte en el mensaje y se extraiga del mismo, independientemente de la naturaleza del medio de transmisión utilizado.

1.3.8. Facilidad criptológica

La demanda de seguridad en la red telefónica convencional no constituye un requerimiento fundamental. Sin embargo, la utilización de servicios telemáticos requiere, en muchas aplicaciones, privacidades más herméticas. Como ya se mencionó, en las aplicaciones de índole militar este requerimiento es fundamental tanto para el servicio vocal como para el de datos.

Por criptología entendemos el proceso por el que se modifica un mensaje de forma tal que sólo el destinatario autorizado puede entenderlo. El proceso debe ser económico tanto para el remitente como para el destinatario, pero al mismo tiempo debe ser muy difícil de descifrar para un intruso (es decir, oneroso en tiempo y equipamiento).

Existe un sinnúmero de técnicas para cifrar mensajes, pero en el caso digital resulta sencillo modificar ciertos bits del mensaje; esto se logra adicionando cierta secuencia binaria, denominada clave, al mensaje original.

Cuando la clave es un 1 se invierte el bit del mensaje y viceversa. Obviamente, el receptor deberá conocer la secuencia de la clave y realizar a su vez otra adición.

Los sistemas digitales permiten satisfacer con suma facilidad el cifrado de mensajes, por lo que el lector puede consultar una amplia bibliografía sobre sistemas criptográficos usados en telecomunicaciones [3][4].

En comparación, la tarea de cifrado en los sistemas analógicos, si bien técnicamente factible, es mucho más difícil de realizar y generalmente no brinda el mismo grado de hermeticidad.

1.3.9. Facilidad de monitoreo

En general la calidad de los sistemas analógicos no puede evaluarse con exactitud mientras el sistema está en servicio. Así, por ejemplo, en las señales multiplex FDM suelen insertarse diversos pilotos que permiten medir la continuidad de los canales y la normalidad de los niveles. La medición del ruido y la distorsión puede realizarse evaluándolos en un intervalo vacante dentro de la banda o bien en los bordes de la misma. En ningún caso, sin embargo, tenemos una medición directa de la calidad del canal en servicio.

Los sistemas digitales, en cambio, en la medida en que manipulan pulsos de nivel y forma definida permiten una medición más directa e independiente de la naturaleza del tráfico que se cursa. Cualquier desviación en la señal recibida más allá de los valores nominales fijados por el diseñador representará una degradación en la calidad del sistema.

Otro método muy difundido para medir la calidad de los sistemas digitales se basa en la observación de las redundancias que se introducen en la señal siguiendo operaciones lógicas definidas para obtener diversos objetivos.

Uno de estos objetivos puede ser la detección y el control de errores, donde la técnica más simple es la introducción de un bit de paridad. Como se verá en el capítulo correspondiente, la observación de este bit permite a los circuitos lógicos del receptor evaluar rápidamente los errores que se producen en el sistema. Además, veremos cómo suelen introducirse otros códigos para facilitar la transmisión en línea. Estos códigos presentan estructuras perfectamente definidas y violarlos también es una indicación de que la calidad del enlace se ha degradado.

1.4. DESVENTAJAS DE LAS REDES DIGITALES

Igual que al considerar las ventajas, ahora analizaremos las desventajas de las redes digitales, comparándolas con las prestaciones que ofrecen los sistemas analógicos, especialmente para la transmisión de señales vocales. Las principales desventajas de las redes digitales son:

- Mayor requerimiento de ancho de banda
- Necesidad de conversión A/D y D/A
- Necesidad de sincronización de tiempo
- Incompatibilidad con la red analógica existente

1.4.1. Mayor requerimiento de ancho de banda

Una señal vocal digitalizada con un nivel de calidad aceptable requiere, como mínimo, ocho veces más ancho de banda. Como ya mencionamos, y como se verá en detalle, la señal digitalizada provee una mayor inmunidad contra el ruido y la interferencia.

En ciertos tramos de la red, como en la línea del abonado, el mayor requerimiento de ancho de banda no es significativo en la medida en que de todas formas está subutilizada. Sin embargo, en otros tramos de la red, especialmente en los canales interurbanos internacionales, el espectro está muy constreñido, y cada intervalo de ancho de banda resulta oneroso. Al mismo tiempo algunos medios de transmisión tienen muy bajos niveles de interferencia y/o ruido, con lo cual la ventaja apuntada de intercambiar el ancho de banda por una mayor inmunidad no es atractiva.

Debe quedar claro que cuando consideramos aisladamente ciertos sistemas de transmisión probablemente resulten más favorables las implementaciones analógicas. Sin embargo, cuando analizamos la red en su conjunto, y que es lo que en general debe hacerse, las ventajas que se obtienen por los sistemas digitales compensan con creces ciertas desventajas localizadas.

Como veremos al tratar la modulación digital, hay algunas técnicas para reducir los anchos de banda de transmisión de las señales digitales a expensas, obviamente, de una menor inmunidad. Esto permite adecuar la transmisión al vínculo elegido, optimando la utilización del ancho de banda y manteniendo la digitalización global de la red.

1.4.2. Necesidad de conversión A/D y D/A

Debe tenerse en cuenta que la mayoría de los sistemas digitales se instalan en un ambiente analógico. Consecuentemente, hay un costo adicional representado por los procesos de conversión de analógico a digital y viceversa.

Sin embargo ciertas consideraciones tienden a atemperar estas desventajas:

- A medida que la red se va convirtiendo en digital, el número de conversiones disminuye, con lo que el costo se reduce, así como las degradaciones que las mismas introducen.
Por ello, se aconseja un grado mínimo de penetración digital en la red para no pagar costos excesivos. Ese grado de penetración depende de cada red en particular, de su configuración, de los costos relativos, etcétera.
- Debe notarse también que los costos electrónicos disminuyen día a día, mientras que otros costos como el cableado, el mantenimiento, etc., cada vez son más significativos. En consecuencia, los costos de conversión son cada vez menores en comparación con el costo total.

1.4.3. Necesidad de sincronización de tiempo

La señal digital que se transmite en una red de servicios integrados no es más que un flujo de dígitos, normalmente binarios. Cuando este flujo llega a un receptor éste debe, entre otras cosas:

- Decidir cuál es el instante correcto para efectuar el muestreo de cada dígito, a efecto de reconocerlos correctamente.
- Identificar el comienzo de las tramas en las que han sido entrelazados los distintos canales y servicios para poder separar y distribuir la información.

Para realizar correctamente estas operaciones, el receptor debe contar con un reloj que le indique exactamente la sucesión de instantes en los que debe llevarlos a cabo. Esta sucesión de instantes es lo que comúnmente se denomina temporización y debe estar perfectamente sincronizada con el flujo de pulsos recibidos.

En general, la generación de la temporización local no ofrece mayores dificultades, por lo que a lo largo de la obra describiremos las consideraciones que deben tenerse en cuenta, así como diversos métodos para obtenerla.

Ahora bien, cuando deseamos interconectar sistemas de transmisión y centrales de conmutación digital con el fin de conformar una red, el problema de la sincronización se torna más complejo. Ya no basta la correcta sincronización de cada sistema individual sino que es necesario establecer un método de sincronización global para toda la red, con el objetivo de que todos los relojes de las centrales digitales operen a igual frecuencia y fase. De esta forma se trata de prevenir deslizamientos en la información digital que se transmite. Estos deslizamientos son los responsables de diversas degradaciones, como pérdida de trama, conexiones cruzadas, desconexiones, etc. Debe aclararse que en general es sumamente difícil y, en consecuencia, muy oneroso, evitar los deslizamientos, por lo cual suele establecerse una tasa máxima de las mismas como objetivo de calidad de la red.

Para concluir, es conveniente puntualizar que si bien en los sistemas analógicos también hay necesidades de sincronización, éstos se deben al sistema de transmisión y más precisamente al tipo de modulación elegido. Así por ejemplo, en los sistemas multiplex FDM que utilizan modulación de banda lateral única con portadora suprimida se requiere la sincronización en frecuencia del oscilador local del receptor para minimizar el error de frecuencia restituida. Obsérvese, sin embargo, que este requerimiento, a diferencia del caso digital, no es inherente a la estructura de la señal.

1.4.4. Incompatibilidad con la red analógica existente

El principal problema para establecer una red totalmente digital es la existencia de una red analógica. Esta dificultad es mayor cuanto más moderna y eficiente es esa red, de allí que en algunos casos los países en desarrollo estén en mejores condiciones de digitalizar la red que las naciones industrializadas. Nadie duda que en poco más de un par de décadas la mayoría de las redes serán en su casi totalidad digitales; sin embargo las estrategias para llegar a ella difieren para cada administración.

Como ya dijimos, las mayores ventajas de una red digital se obtienen cuando está totalmente digitalizada. Ello significa que la digitalización debe comenzar en el domicilio del abonado, con lo que no sólo mejoraría la calidad total del servicio sino que se obtendrían ventajas adicionales como la eliminación de interfaces y de ecos, facilidades de multiplexado, amén de la mencionada posibilidad de acceder a una vasta diversidad de servicios.

No obstante estas ventajas evidentes, la inmediata transición al lazo de abonado digital demandaría un nivel de inversión tan elevado que lo torna prácticamente inaplicable.

Una de las grandes preocupaciones de las empresas de telecomunicaciones es el mantenimiento de la diversidad de equipos que componen la red. La coexistencia de equipamientos de distintas generaciones, todos ellos en condiciones de operatividad, demandan multiplicación de repuestos, variación de correctores, diversidad de valores nominales de uso, etc. La introducción de nuevos equipamientos digitales en el corto plazo aumenta esa diversidad, y lo que es peor, al tratarse de concepciones totalmente diferentes requiere intensos programas de capacitación del personal, nuevos instrumentos de medición, un nuevo almacén de repuestos, etcétera.

Por supuesto que hay ventajas de índole económico que tienden a compensar estos inconvenientes, como por ejemplo: reducción del costo comparativo de los equipos de tecnología digital, menos espacio operativo, ya sea en equipos en los distribuidores de cable, aprovechamiento intensivo de la planta exterior, etcétera.

Por todo ello, la estrategia de introducción de la digitalización de la red es específica para cada administración y a ella concurren un sinnúmero de variables, las que no son exclusivas del área de telecomunicaciones.

I.5. CONCLUSIÓN

En los próximos capítulos consideraremos los diversos aspectos que integran el funcionamiento de las redes de comunicación digitales.

No queremos dejar de reflexionar que la tecnología de las telecomunicaciones está evolucionando de forma tal que tendrá un profundo impacto sobre toda la sociedad, tanto en los países en desarrollo como en los desarrollados. Las nuevas redes probablemente cambiarán los esquemas laborales, la educación, la industria, el cuidado de la salud y también nuestro tiempo libre. Su influencia puede ser decisiva en los medios masivos de comunicación, en los procesos de gobierno, e incluso podemos esperar que contribuya a fortalecer una democracia más participativa.

Las consecuencias políticas y sociales de las nuevas redes digitales de comunicación no pueden ser ignoradas por nadie, pero mucho menos por los hombres que las proyectan, construyen y operan.

BIBLIOGRAFÍA

- [1] Shannon, C. E., *Communication in the Presence of Noise*, *Proc. IRE* vol. 37, enero de 1949.
- [2] Still, A., *Communication Through the Ages*, Nueva York, Murray Hill Books, 1946.
- [3] Roden, Martin, *Digital and Data Communication Systems*, Prentice-Hall, 1982.
- [4] Diffie, W. y M. E. Hellman, *New Directions in Cryptography*, *IEEE Trans. Inf. Theory*, vol. I, t. 22, núm. 6, 1976.

CAPÍTULO II

TEORÍA DE LA INFORMACIÓN

II.1. INTRODUCCIÓN

LAS FRASES TEORÍA DE LA COMUNICACIÓN y teoría de la información designan lo mismo. Aquí empleamos información como un término técnico que no debe confundirse con su interpretación convencional. En comunicaciones, información es simplemente lo que produce la fuente para transferirla al usuario. Esto significa que antes de la transmisión, la información no está disponible en el destinatario; de otro modo la transferencia sería cero.

La teoría de la comunicación nació con la telegrafía ante la necesidad de especificar precisamente “la capacidad” de varios sistemas de telecomunicaciones para llevar información. El primer ensayo para formular una medida matemática lo hizo Hartley en 1928, y sus ideas son las bases de la teoría actual. A los ingenieros les interesa la transmisión correcta de señales o la representación eléctrica de mensajes sin importales por lo general el objetivo de estos mensajes. Arbitrando los medios para que el telégrafo o el teléfono transmita sus señales con fidelidad, los mensajes tendrán significado, valor, certeza, confiabilidad y todas sus otras propiedades. Las señales deben ser correctas, y entonces todas estas propiedades humanas son inherentes y consecuentes. La teoría matemática de la información se centra solamente en las señales y su contenido de información, y se abstrae de todos los usos humanos específicos.

La información sólo puede recibirse cuando hay duda; y las dudas implican existencia de alternativas, lo cual trae aparejado elección, discriminación o selección.

Pero la selección (o discriminación) puede llevarse a cabo en enlaces de comunicaciones no humanos. Por ejemplo, en un teletipo se oprime una tecla cada vez y las señales eléctricas codificadas que llegan al receptor seleccionan y oprimen las teclas correctas automáticamente. Nos da la sensación de que las teclas del receptor están siendo presionadas por dedos invisibles.

Esas señales eléctricas codificadas que llegan al receptor llevan cierto contenido de información en virtud de su potencial para decidir elecciones. Las señales son opciones para el receptor, quien habrá de elegir las de un conjunto de distintos signos al que denominamos *alfabeto* y que bien pueden ser letras de un lenguaje escrito, números, palabras impresas, ordenadas de una forma de onda; en fin, cualquier tipo de signos creados para la comunicación.

Peró ese alfabeto debe especificarse antes de que el contenido de información de los mensajes pueda discutirse numéricamente, y además debe suponerse que el mismo alfabeto exista en el terminal transmisor y receptor de un canal de comunicación. Así, la función de la fuente de información es seleccionar signos sucesivamente de este alfabeto, constituir mensajes, y transmitirlos en forma física como señales por medio de un canal al receptor. En el receptor las señales operarán un alfabeto idéntico y seleccionarán los signos correspondientes, o sea, los mensajes enviados y recibidos.

Hay que notar la diferencia entre mensaje y señal. Un mensaje se considera como las elecciones del alfabeto, puesto en forma física (señales) para la transmisión.

Claude Shannon, en su trabajo sobre la teoría matemática de la comunicación, se planteó un objetivo básico: hacer lo más eficiente posible la transmisión de información de un punto a otro. Por ejemplo, transmitir mensajes lo más rápidamente posible y con el número mínimo de errores. Hay un límite en la tasa de información que puede ser transmitido por ese sistema; este límite es la capacidad de información, la cual viene determinada por las limitaciones físicas fundamentales en la transmisión de la información por medios eléctricos, como el ancho de banda y el ruido. Es importante recordar que por medio de técnicas de modulación y codificación es posible intercambiar ancho de banda por relación señal a ruido.

Shannon, para encarar lo anterior, se planteó lo siguiente: dado un conjunto de posibles mensajes que una fuente puede transmitir no de nuestra elección, ¿cómo pueden representarse éstos de la mejor manera posible para llevar información sobre un sistema con sus limitaciones físicas inherentes? Para tratar este problema es necesario concentrarse más sobre la *información* que sobre las señales, y por esta razón el trabajo de Shannon fue rebautizado como *teoría de la información*.

La teoría matemática de la información trata con tres conceptos básicos: la medida de la información, la capacidad de un canal de comunicación para transferir información y la codificación como un medio para utilizar los canales a su completa capacidad. Estos conceptos pueden enlazarse en el teorema fundamental de la teoría de la información:

Dada una fuente de información y un canal de comunicación, existe una técnica de codificación tal que la información puede ser transmitida sobre el canal con una tasa menor que la capacidad del canal y con una frecuencia arbitrariamente pequeña de errores, a pesar de la presencia de ruido.

Lo sorprendente de este teorema es la transmisión casi libre de errores sobre un canal ruidoso, que se logra por medio del uso de la codificación. En esencia, la codificación se usa para adaptar la fuente y el canal para máxima transferencia de información confiable, algo casi análogo a la adaptación de impedancia para máxima transferencia de potencia. En este capítulo trataremos de contestar los siguientes interrogantes:

- ¿Cómo se mide la cantidad de información?
- ¿Cómo se mide la capacidad de un canal de comunicación?
- ¿La acción del transmisor en cambiar el mensaje a señal representa a menudo un proceso de codificación?
- ¿Cuáles son las características eficientes de un proceso de codificación?, y cuando la codificación es tan eficiente como sea posible, ¿a qué tasa puede llevar información el canal?
- ¿Cómo afecta el ruido la exactitud del mensaje recibido? ¿Cómo se pueden minimizar los efectos indeseables de ruido?
- Si la señal transmitida es analógica (voz o música) en lugar de símbolos discretos (telegrafía o palabra escrita), ¿cómo afecta al problema?

Podemos adelantar que la teoría desarrollada para el caso discreto se puede extender al caso de señales continuas (analógicas), y aunque es más difícil y complicada matemáticamente no es en esencia diferente. Además, cualquier señal analógica limitada en banda puede digitalizarse por medio de un proceso de muestreo, cuantificación y codificación.

II.2. MEDIDA DE LA INFORMACIÓN

Si examinamos con detalle el contenido de información de un mensaje podremos ahorrar esfuerzo en su transmisión de un punto a otro. Por ejemplo, si deseamos enviar un telegrama de felicitación por un casamiento o nacimiento, la compañía de telégrafos enviará un texto estandarizado, típico. El operador simplemente escribirá el destino, el remitente y un número determinado que identifique el texto modelo.

En un restaurante que ofrece la posibilidad de elegir entre seis almuerzos completos, no es necesario leer los platos del almuerzo elegido, bastará especificar el número. En un sentido instintivo podemos ver que algunos mensajes largos no contienen gran información (por ejemplo al leer uno de los almuerzos en lugar de especificar el número).

El concepto de contenido de información en un mensaje particular debemos formalizarlo para luego encontrar que cuanto menor es la información en un mensaje, más rápido podemos transmitirlo.

El concepto de información está ligado al de predictibilidad: cuanto más probable es un mensaje menor es la información transmitida por ese mensaje. Por ejemplo, si en el lugar de trabajo cobramos el sueldo indefectiblemente el último día hábil del mes, llamar a un compañero para comunicarle esta novedad no le transmitirá gran información, pues la probabilidad que ocurra es prácticamente 1. Pero si le comunico que la compañía ha decidido pagar un premio extra de siete salarios, la cantidad de información es distinta, pues la probabilidad es mucho menor que en el caso anterior.

Por lo tanto, podemos decir que la medida de la información se relaciona con la incertidumbre que el usuario tiene acerca de lo que será el mensaje. Pero hay que destacar que la cantidad de información depende más de la incertidumbre del mensaje que de su contenido o posibles interpretaciones.

Yendo al otro extremo del sistema de comunicación, la medida de la información indica la libertad de elección que puede tener la fuente en seleccionar un mensaje. Si la fuente puede elegir libremente entre varios mensajes diferentes, el usuario tiene una alta incertidumbre respecto al mensaje que será seleccionado; pero si no hay elección, no hay incertidumbre y por lo tanto información.

Es evidente que la medida de la información comprende probabilidades. Los mensajes de alta probabilidad indican poca incertidumbre del usuario o poca elección en la fuente, llevando una pequeña cantidad de información, y viceversa. Este concepto lo formalizaremos definiendo la información mutua en términos de probabilidad.

II.3. INFORMACIÓN MUTUA

Consideremos una fuente que produce varios mensajes. Sea uno de los mensajes A y P_A la probabilidad de que sea elegido para su transmisión. De acuerdo con lo ya expresado, escribimos la información mutua asociada con A como:

$$I_A = f(P_A)$$

donde la función $f(P_A)$ debe ser determinada. Para encontrar $f(P_A)$ es intuitivo suponer los siguientes requerimientos:

$$f(P_A) \geq 0 \quad \text{donde} \quad 0 \leq P_A \leq 1 \quad (\text{II.1})$$

$$\lim_{P_A \rightarrow 1} f(P_A) = 0 \quad (\text{II.2})$$

$$f(P_A) > f(P_B) \quad \text{para } P_A < P_B \quad (\text{II.3})$$

Muchas funciones satisfacen las ecuaciones (II.1) a (II.3), pero la decisión final se obtiene al considerar la transmisión de mensajes independientes. Cuando el mensaje A es entregado al usuario recibe I_A unidades de información. Si un segundo mensaje es entregado, la información total recibida debería ser la suma de las informaciones mutuas $I_A + I_B$.

Esto es aún más claro si consideramos que A y B vienen de diferentes fuentes. Pero supongamos que A y B vengan de la misma fuente: entonces podemos hablar del mensaje compuesto $C = A \cdot B$. Si A y B son estadísticamente independientes,

$$P_C = P_A \cdot P_B \quad I_C = f(P_A \cdot P_B)$$

Pero la información recibida es aún

$$I_C = I_A + I_B = f(P_A) + f(P_B) \quad \text{y así} \quad (\text{II.4})$$

$$f(P_A \cdot P_B) = f(P_A) + f(P_B)$$

que es nuestro requerimiento para $f(P_C)$.

Hay una sola función que satisface las condiciones (II.1) a (II.4) y es la función logarítmica $f(x) = \log_b(x)$ donde b es la base del logaritmo. Así, la información mutua se define como:

$$I_A = \log_b P_A = \log_b \frac{1}{P_A} \quad (\text{II.5a})$$

Como $0 \leq P_A \leq 1$ el logaritmo es positivo, como se desea.

Especificando ahora la base b del logaritmo seleccionamos la unidad de información. La más usada es con base 2 y su unidad de información se denomina bit.

$$I_A = \log_2 \frac{1}{P_A} \text{ [bit]} \quad (\text{II.5b})$$

Si $P_A = P_B = 1/2$ entonces $I_A = I_B = \log_2 2 = 1$ bit, que es la cantidad requerida o llevada por la elección entre dos sucesos igualmente probables (equiprobables).

Puesto que las tablas de logaritmos en base 2 no son comunes, es útil recordar las siguientes relaciones:

$$\log_2 x = \log_2 10 \cdot \log_{10} x \approx 3.32 \log_{10} x \quad (\text{II.5c})$$

También se puede utilizar

$$\log_b x = \frac{\log x}{\log b} = \frac{\ln x}{\ln b}$$

Si hubiéramos tomado como base b en la ecuación (II.5) de los logaritmos naturales $e = 2.71828$, tendríamos

$$I(A) = \ln \frac{1}{P(A)} \quad \text{Nats}$$

tomando logaritmos decimales,

$$I(A) = \log_{10} \frac{1}{P(A)} \quad \text{Hartley}$$

Es evidente que: 1 Hartley = 3.32 bits y que 1 Nat = 1.44 bits, recordando que:

$$\ln_e x = 2.30258 \cdot \log_{10} x = \frac{2.30358 \times \log_2 x}{3.32} = 0.6935 \cdot \log_2 x = \frac{1}{1.44} \cdot \log_2 x$$

En este capítulo, si no se especifica lo contrario, todos los logaritmos serán de base 2.

II.4. BINIT Y BIT

Es interesante observar que la palabra dígito binario (*binary digit*), abreviada bit, se debe a que dos estados pueden ser presentados por los dígitos binarios: 0 y 1. Pero un dígito binario podría llevar más de un bit de información o menos, dependiendo de su probabilidad de ocurrencia. Por eso no es siempre correcto decir que un 1 o un 0 es un bit, pues esto se puede interpretar como su unidad de información (sólo es cierto para sucesos equiprobables). Por ejemplo, si

$$P_A = 1/4 \quad \text{y} \quad P_B = 3/4$$

entonces

$$I_A = \log_2 4 = 2 \text{ bits} \quad \text{y} \quad I_B = \log_2 4/3 = 0.414 \text{ bits}$$

Para evitar errores de interpretación, a los dígitos binarios como elementos de mensajes se les llama binits en lugar de bits (denominación habitual), para no confundirlos con su información asociada.

Por ejemplo, un tren de pulsos eléctricos como el de la figura II.1 está compuesto de binits 1 y binits 0:

La aparición de cualquier binit 1 da una información de $\log_2 3 = 1.58$ bits.

La aparición de cualquier binit 0 da una información de $\log_2 3/2 = 0.58$ bits.

Esta aclaración es para evitar que a un dígito del tren de pulsos lo llamemos 1 bit (contracción de dígito binario), pues podría interpretarse como unidad de información, cuando en realidad ésta depende de la probabilidad.

Como en la práctica es frecuente seguir llamando erróneamente bits a los dígitos binarios (binits), otra alternativa propuesta para desechar esto definitivamente es mantener la denominación bits para los dígitos binarios (binits) y llamar Shannon a la unidad de información.

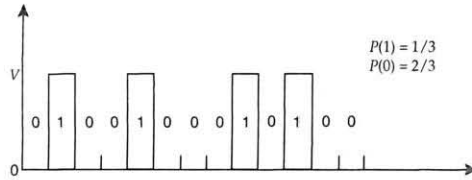


Figura II.1. Tren de pulsos binarios

Ejemplo II.1

a) Calcule la información asociada a la caída de una moneda (suceso estadísticamente independiente).

Respuesta: 1 bit

b) Calcule la información entregada por la aparición de una letra entre 32 equiprobables.

Respuesta: 5 bits

Ejemplo II.2

Supongamos que una fuente produce los símbolos A , B , C y D con probabilidades $1/2$, $1/4$, $1/8$ y $1/8$.

a) Calcule la información en cada caso.

b) Si los símbolos son independientes, calcule los bits de información del mensaje BACCA.

Respuesta a): 1; 2; 3; 3 bits

Respuesta b): $2 + 1 + 3 + 1 = 7$ bits

Ejemplo II.3

Calcule la probabilidad de que aparezcan tres caras consecutivas en la tirada de una moneda y su información asociada.

Respuesta:

$$P_{ccc} = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8} \qquad I_{ccc} = \log_2 8 = 3 \text{ bits}$$

Ejemplo II.4

De un mazo de 40 barajas españolas se extrae una carta. Si me informan que es de oros, ¿cuántos bits de información he recibido?, ¿qué información adicional se requiere para especificar la carta?

Respuesta:

$$I_1 = \log_2 4 = 2 \text{ bits} \qquad I_2 = \log_2 10 = 3.32 \text{ bits}$$

Ejemplo II.5

Hay un proverbio que dice que una imagen vale más que mil palabras. Verifiquémoslo suponiendo que descomponemos el cuadro en un número de puntos discretos o elementos; cada uno tiene un nivel de brillo que varía en pasos desde el negro hasta el blanco.

Respuesta:

Para el análisis supongamos una imagen compuesta de 400 líneas horizontales con 300 puntos discretos por línea horizontal, y ocho niveles equiprobables fácilmente distinguibles de brillo; además supongamos un vocabulario de cien mil palabras equiprobables. Hay $8^{1.2 \cdot 10^5}$ posibles imágenes, cada una con una probabilidad

$$P = \frac{1}{8^{1.2 \cdot 10^5}}$$

y una información $I = \log_2 1/P = 360\,000$ bits.

A este mismo resultado podríamos haber llegado suponiendo que la información por punto o elemento es $\log_2 8 = 3$ bits, lo que da una información total para la imagen de $3 \times 120\,000 = 360\,000$ bits.

Respecto al vocabulario cada palabra tiene una $I = \log_2 10^5 = 16.6$ bits y 1 000 palabras tienen una información de $16.6 \times 1\,000 = 16\,600$ bits.

Luego, para el vocabulario y la imagen supuestos se cumple el proverbio que dice que una imagen tiene mayor información que mil palabras. La validez de las suposiciones no tiene sentido discutirla, pues aquí nos interesa el método, no los resultados.

II.5. INFORMACIÓN PROMEDIO (ENTROPÍA)

Definimos la información mutua en función de los mensajes individuales o símbolos que una fuente puede producir. Pero ésta no es una descripción útil de la fuente en lo que se refiere a los sistemas de comunicaciones, ya que no se diseña para un mensaje en particular sino para todos los mensajes posibles. Por lo tanto, aunque el flujo de información instantáneo de una fuente puede ser errático, debemos describir la fuente en términos de la información promedio producida. Esta información promedio también se llama entropía de la fuente.

Para una fuente discreta cuyos símbolos son estadísticamente independientes, la expresión entropía es fácil de formular. Sea m el número de símbolos diferentes, es decir, un alfabeto de longitud m . Cuando el símbolo j se transmite, éste lleva $I_j = \log 1/P_j$ bits de información. En un

mensaje largo de $N \gg$ símbolos, el símbolo j ocurre NP_j veces, y la información total en el mensaje es aproximadamente

$$NP_1 I_1 + NP_2 I_2 + NP_3 I_3 + \dots + NP_m I_m = \sum_{j=1}^m NP_j I_j \text{ bits} \quad (\text{II.6})$$

La que dividida por N da la información promedio por símbolo. Por lo tanto, definimos la entropía de una fuente discreta como

$$H = \sum_{j=1}^m P_j I_j = \sum_{j=1}^m P_j \cdot \log_2 1/P_j \text{ (bits/símbolo)} \quad (\text{II.7})$$

Nótese que la ecuación (II.7) es un promedio del ensamble. Si la fuente no es estacionaria, las probabilidades de los símbolos pueden cambiar con el tiempo y la entropía no es significativa. Nosotros supondremos que las fuentes de información son ergódicas, de tal modo que los promedios del ensamble y del tiempo son idénticos.

La entropía nos indica que en promedio nosotros esperamos obtener: H bits de información por símbolo o NH bits en un mensaje de N símbolos, si N es grande.

Para un alfabeto de longitud fija (m fija) la entropía de una fuente discreta depende de las probabilidades de los símbolos, pero es limitada por

$$0 \leq H \leq \log_2 m \quad (\text{II.8})$$

El límite inferior $H = 0$ implica que la fuente no entrega información (en promedio), así que aquí no hay incertidumbre en el mensaje. Esto correspondería a una fuente que tiene un símbolo con $P = 1$ y todos los demás símbolos con probabilidad cero.

La entropía máxima $H = \log_2 m$ corresponde al máximo de incertidumbre o máxima libertad de elección. Esto implica que todos los símbolos son igualmente probables, o sea, no hay un símbolo preferido. Entonces $H = H_{\max} = \log_2 m$ cuando todos los símbolos tienen una probabilidad $P_j = 1/m$ (ejemplo II.6).

La variación de H entre los límites de la ecuación (II.8) se observa considerando una fuente binaria ($m = 2$) con probabilidades p y $q = 1 - p$; graficando

$$H = p \log \frac{1}{p} + (1 - p) \cdot \log \frac{1}{1 - p} \quad (\text{II.9})$$

Como se observa en la figura II.2, la máxima entropía $H = \log_2 2 = 1$ bit ocurre para $p = 0.5$, o sea el caso equiprobable.

Ejemplo II.6

Una fuente discreta tiene salidas x_j para $1 \leq j \leq N$ con probabilidades $p(x_j)$. Demuestre que la entropía de la fuente tiene un valor máximo de $\log_2 N$ y que el mismo ocurre cuando todas las salidas de la fuente son equiprobables.

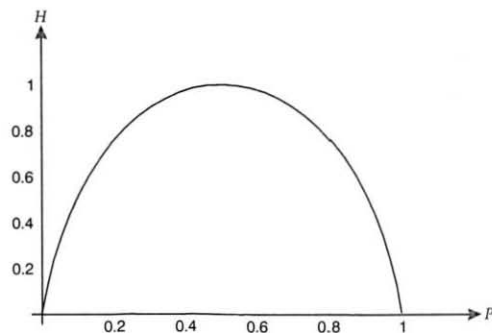


Figura II.2. Entropía de una fuente binaria en función de la probabilidad de un símbolo

Respuesta: Una manera es variar $p(x_k)$ y $p(x_n)$ mientras mantenemos las otras probabilidades, $p(x_j)$, constante para $j \neq 1$ o N . Entonces

$$p(x_k) = 1 - \sum_{\substack{j=1 \\ j \neq k}}^N p(x_j)$$

Ahora determinaremos el valor de $p(x_k)$ que maximiza $H(x)$. Diferenciando

$$\begin{aligned} \frac{dH}{dp(x_k)} &= -\frac{d}{dp(x_k)} \left[p(x_k) \log_2 p(x_k) + p(x_n) \log_2 p(x_n) \right] \\ \frac{dH}{dp(x_k)} &= - \left[\log_2 p(x_k) + p(x_k) \frac{1}{p(x_k)} \log_2 e + (-1) \log_2 p(x_n) + p(x_n) \frac{1}{p(x_n)} \log_2 e (-1) \right] \\ &= -\log_2 p(x_k) + \log_2 p(x_n) = \log_2 \frac{p(x_n)}{p(x_k)} \end{aligned}$$

Colocando la derivada igual a cero tenemos:

$$\log_2 \frac{p(x_n)}{p(x_k)} = 0$$

Así, $p(x_n) = p(x_k)$

Puesto que x_k es arbitraria, tenemos:

$$p(x_k) = p(x_N), \quad i = 1, 2, \dots, N-1$$

con lo que $P(x_k) = 1/N$ para todo N .

Por lo tanto, el máximo valor de entropía se logra cuando todas las salidas tienen igual probabilidad. Este valor es:

$$H(x) = \sum_{j=1}^N \frac{1}{N} \log_2 N = \log_2 N$$

Cabe destacar que nosotros sabemos que maximizamos la entropía puesto que el valor mínimo de la entropía es 0. Este mínimo se logra cuando una salida tiene probabilidad uno y todas las otras tienen probabilidad cero.

II.6. TASA DE INFORMACIÓN

Supongamos que dos fuentes tienen la misma entropía pero una es más rápida que la otra, es decir, produce más símbolos por unidad de tiempo.

En un periodo dado, más información se transfiere de la fuente más rápida, lo que causa necesidades mayores sobre el sistema de comunicación. Por lo tanto, la descripción de una fuente no es por su entropía solamente, sino por su tasa de información media en bits por segundo. La tasa de información de una fuente discreta se define como

$$R = \frac{H}{\bar{\sigma}} \text{ bits/seg} \quad (\text{II.10})$$

donde $\bar{\sigma}$ es la duración promedio del símbolo.

Por lo tanto $1/\bar{\sigma}$ es igual al número promedio de símbolos por unidad de tiempo que denominaremos (velocidad de señalización) y cuya unidad es binit/seg = baudio. A la velocidad de señalización la indicaremos con la letra r .

Ejemplo II.7

Calcule la tasa de información de una fuente telegráfica teniendo por probabilidades y duración promedio:

$$P \text{ punto} = 2/3 \quad P \text{ raya} = 1/3$$

$$\sigma \text{ punto} = 0.2 \text{ seg} \quad \sigma \text{ raya} = 0.4 \text{ seg}$$

Respuesta:

$$H = 2/3 \log_2 3 + 1/3 \log_2 3 = 0.920 \text{ bit/símbolo}$$

$$\bar{\sigma} = 2/3 \times 0.2 + 1/3 \times 0.4 = 0.267 \text{ seg}$$

$$R = \frac{0.920}{0.267} = 3.44 \text{ bits/seg.}$$

Ejemplo II.8

Una fuente produce cinco símbolos con probabilidades, $1/2$; $1/4$; $1/8$; $1/16$ y $1/16$. Calcule H :

Respuesta:

$$\begin{aligned} H &= \frac{1}{2} \log_2 2 + \frac{1}{4} \log_2 4 + \frac{1}{8} \log_2 8 + \frac{1}{16} \log_2 16 + \frac{1}{16} \log_2 16 \\ &= 0.5 + 0.5 + 0.375 + 0.25 + 0.25 = 1.875 \text{ bits/símbolo} \end{aligned}$$

Ejemplo II.9

Una fuente de datos tiene ocho símbolos equiprobables y emite en blocks de tres a una tasa de 1 000 blocks por segundo. Si el primer símbolo de cada block es siempre el mismo (sincronización) y los dos restantes pueden ser cualquiera de los ocho símbolos de la fuente, calcule la tasa de información R .

Respuesta:

$$\begin{aligned} H_{\text{símbolo}} &= \log_2 m = \log_2 8 = 3 \text{ bits/símbolo} \\ H_{\text{block}} &= (H_{\text{símbolo}}) (\# \text{ símbolos por block}) = 3 \cdot 2 = 6 \text{ bits/block} \\ \bar{\sigma}_{\text{block}} &= \frac{1}{1\,000} \text{ seg/block} \\ R &= \frac{H_{\text{block}}}{\bar{\sigma}_{\text{block}}} = \frac{6 \text{ bit/block}}{\frac{1}{1\,000} \text{ seg/block}} = 6\,000 \text{ bits/block} \end{aligned}$$

Ejemplo II.10

Se envía un mensaje utilizando cinco pulsos de igual duración. El primer pulso se restringe a dos niveles mientras que los otros cuatro pueden tomar cualquier nivel de entre $+4$ y -4 , incluido el cero.

- ¿Cuántos mensajes se pueden enviar en un grupo simple de cinco pulsos?
- Si el mensaje se transmite como una secuencia de pulsos binarios, ¿cuántos pulsos serían necesarios?
- Calcule la relación de anchos de banda si se considera que el intervalo total del tren de pulsos de $a)$ y $b)$ son iguales.

Respuesta:

- 13 122 mensajes = $2^1 \cdot 9^4$
- $14; 2^{14} \geq 13\,122$
- $14/5 = 2.8$

NOTA: Es importante observar que en el caso *a*) se puede enviar la misma información que en el caso *b*), pero en un ancho de banda 2.8 veces menor. Si se desea que la diferencia entre dos niveles contiguos sea la misma para ambos casos, a efecto de tener la misma facilidad en la detección en el caso de ruido gaussiano aditivo, por ejemplo, evidentemente en el caso *a*) se debe transmitir mayor potencia que en el caso *b*).

O sea, la reducción de ancho de banda en el caso *a*) se paga con potencia si se quiere conservar la misma facilidad de detección que en el caso *b*) para una misma transmisión de información.

Ejemplo II.11

Una fuente de datos tiene 16 símbolos equiprobables posibles, cada uno de 1 mseg de duración. Los símbolos se emiten en bloques de 15, separados por un pulso de sincronización de 5 mseg. Calcule *R*.

Respuesta:

$$H_{\text{block}} = (H_{\text{símbolo}}) (\text{Núm. de símbolos por block}) = (\log_2 16) (15)$$

$$H_{\text{block}} = 4 \times 15 = 60 \text{ bit/block}$$

$$\bar{\sigma}_{\text{block}} = 0.02 \text{ seg/block}$$

$$R = \frac{H_{\text{block}}}{\bar{\sigma}} = \frac{60}{0.02} = 3\,000 \text{ bits/seg}$$

Ejemplo II.12

Un sistema de facsímil transmite una imagen que tiene 250 líneas horizontales y 200 puntos por línea. Si cada punto puede tomar 32 niveles equiprobables de brillo, calcule la información de la imagen.

Si una vez explorada la imagen se almacena eléctricamente en una memoria y se envía por un canal de transmisión a 208.33 bit/seg, ¿cuánto tiempo se tarda en transmitirla?

Respuesta:

$$H_{\text{imagen}} = 5 \cdot (50\,000) = 250\,000 \text{ bits}$$

$$\text{Tiempo de transmisión} = \frac{250\,000}{208.33} = 1\,200 \text{ seg} = 20 \text{ min.}$$

Ejemplo II.13

Una imagen de tv tiene 625 líneas con 500 puntos por línea y 128 niveles equiprobables de brillo por punto. Si se transmiten 25 imágenes por segundo, calcule *R*.

Respuesta:

$$H_{\text{imagen}} = (\log_2 128) (625 \times 500) = 2.1875 \times 10^6$$

$$R = \frac{H_{\text{imagen}}}{\sigma} = (2.1875 \times 10^6) (25) = 54.6875 \text{ Megabits /seg}$$

Ejemplo II.14

Se envía un mensaje usando cinco pulsos cada uno con una duración de 1 mseg y dos niveles 0 y V posibles de tensión (equiprobables). La cantidad de mensajes diferentes que se pueden enviar con este tren de cinco pulsos es $2^5 = 32$ mensajes y la tasa de información es

$$R = 1 \text{ bit/1 mseg} = 1\,000 \text{ bits/seg}$$

La velocidad de señalización es $r = 1\,000$ baudios.

Si los pulsos podrían tomar cada uno cuatro niveles posibles de tensión en forma equiprobable (la misma duración anterior), se podrían enviar $4^5 = 1\,024$ mensajes diferentes, y

$$R = \frac{2 \text{ bit}}{1 \text{ mseg}} = 2000 \frac{\text{bit}}{\text{seg}}, \quad r = 1\,000 \text{ baudios}$$

Vemos que en el primer caso se transmite 1 bit por baudio y en el segundo caso 2 bits por baudio, ¿cuál es el costo de esta ventaja?

Si se quiere que el nivel máximo de tensión transmitido en el segundo caso siga siendo V como en el primer caso, la mejora de 2 bits por baudio hará que los niveles de decisión en el proceso de detección (para ruido gaussiano aditivo, por ejemplo) estén más juntos entre sí y se obtenga una mayor probabilidad de error en la detección de niveles.

Si se quiere obtener la misma probabilidad de error que para el primer caso es necesario aumentar los niveles de tensión en el segundo caso (aumento de potencia).

Una manera de lograr transmitir 1 024 mensajes sin modificar los dos niveles posibles de tensión del primer caso es transmitiendo 10 pulsos binarios en el intervalo de 5 mseg ($2^{10} = 1\,024$ mensajes); pero atención, R es 2 000 bits/seg, pero tengo 1 bit por baudio pues se tuvo que duplicar la velocidad de señalización al doble. De lo analizado vemos que hay un intercambio entre la tasa de información transmitida en (bits/seg) y el ancho de banda y la relación señal ruido.

Ejemplo II.15

Se quieren transmitir pulsos de K niveles de tensión equiprobables por un canal telefónico. La velocidad de señalización en baudios que se puede transmitir por un canal o vínculo depende del ancho de banda B . Por ejemplo, para un canal telefónico una velocidad de señalización típica es de 2 400 baudios. Pero la cantidad de bits por baudio que se transmite depende de la cantidad K de niveles que se elija.

Si $K = 2$, se tiene 1 bit por baudio $\therefore 2\,400 \text{ bits/seg}$

Si $K = 4$, se tienen 2 bits por baudio $\therefore 4\,800 \text{ bits/seg}$

Si $K = 8$, se tienen 3 bits por baudio $\therefore 7\ 200$ bits/seg

Si $K = 16$, se tienen 4 bits por baudio $\therefore 9\ 600$ bits/seg

Pero en los cuatro casos la velocidad de señalización es de 2 400 baudios (depende del ancho de banda B del canal).

Ejemplo II.16

Una máquina teletipadora da una vuelta cada 150 mseg, produciendo un carácter o letra de acuerdo con el código de la figura II.3. El primer pulso (arranque) y el último pulso (parada) se utilizan para sincronización y los cinco restantes para información.

Una misma codificación puede significar dos caracteres distintos, dependiendo de si se le antepusieron cifras o letras.

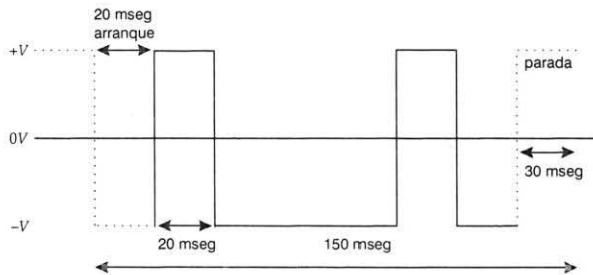


Figura II.3. Señal telegráfica

Determine cuántas palabras por minuto se pueden transmitir si en promedio se tienen seis caracteres por palabra (incluido el espacio).

Respuesta:

150 mseg: 1 carácter

60 000 mseg: 400 caracteres

Luego 400 caracteres: $400/6 = 66.66$ palabras/minuto

Otra manera de llegar a la misma respuesta es la siguiente: para símbolos equiprobables tengo 5 bits en 150 mseg, o sea $33 \text{ bits/seg} = 33 \times 60 \text{ bits/minuto}$. La información de los pulsos de sincronización (arranque y parada) es cero, puesto que su probabilidad de ocurrencias es igual a 1.

La cantidad de palabras por minuto es

$$\frac{1\ 980 \text{ bits/minuto}}{5 \text{ bits} \cdot 6 \text{ letras}} = 66.66 \text{ palabras/minuto}$$

Ejemplo II.17

Calcule la velocidad de señalización de la señal telegráfica del ejercicio anterior.

Respuesta:

$$r = \frac{1}{0.02 \text{ seg}} = 50 \text{ baudios}$$

Ejemplo II.18

Se transmiten trenes o bloques de cuatro pulsos cada uno, cada pulso con una duración de 1 mseg. Los pulsos pueden tomar de forma equiprobable los niveles de tensión 0; 1; 2, y 3 V, excepto el primer pulso del tren o bloque que siempre toma el nivel 1 volt. Calcule la tasa de información R .

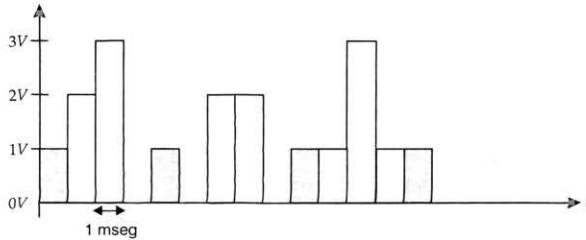


Figura II.4. Trenes de cuatro pulsos cada uno

Respuesta:

$$R = \frac{H}{\sigma} = \frac{(\log_2 4) (3) \text{ bits/block}}{4 \text{ mseg /block}} = 1500 \text{ bits/seg}$$

Ejemplo II.19

Se transmiten pulsos de 1 mseg de duración cuyos niveles pueden ser: 1; 2; 3, o 4 volt, con probabilidades $1/2$, $1/4$, $1/8$ y $1/8$. Calcule R

Respuesta:

$$H = \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 3 + \frac{1}{8} \cdot 3 = 1.75 \text{ bits/binit}$$

$$R = \frac{H}{\sigma} = \frac{1.75}{1 \text{ mseg}} = 1750 \text{ bits/seg} = H \cdot r = 1.75 \frac{\text{bit}}{\text{binit}} \cdot 1000 \frac{\text{binit}}{\text{seg}}$$

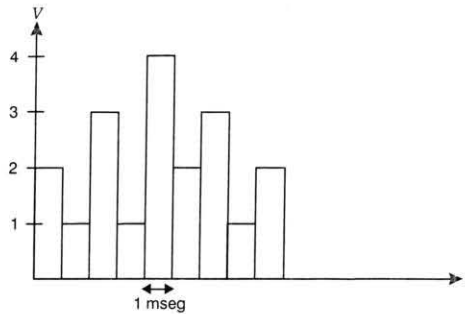


Figura II.5. Pulsos de cuatro niveles

Ejemplo II.20

Al tren de pulsos del ejemplo anterior se le codifica para su transmisión de acuerdo con el siguiente código binario:

Nivel	Probabilidad	Código binario
1v	1/2	00
2v	1/4	01
3v	1/8	10
4v	1/8	11

- ¿Cuál es la duración de un binit a la salida del codificador?
- Calcule cuántos binitos por bit se necesitan para la transmisión.
- Calcule la probabilidad de aparición de un binit 0 y un binit 1 a la salida del codificador. ¿Son equiprobables a los dígitos binarios? ¿Los símbolos binarios son estadísticamente independientes a la salida del codificador?
- Calcule R por medio de la ecuación de entropía para símbolos estadísticamente independientes. ¿Por qué no dio el valor de 1.750 bits/seg hallado previamente?
- ¿Se eligió una codificación adecuada?

Respuestas:

a) 0.5 mseg

b) $\frac{2\,000}{1\,750} = 1.142$ binitos/bit

c) $P(0) = \frac{1}{2} + \frac{1}{8} + \frac{1}{16} = \frac{11}{16}$; $P(1) = \frac{1}{8} + \frac{1}{16} + \frac{1}{8} = \frac{5}{16}$; No; No

d) $H = \frac{11}{16} \cdot \log_2 \frac{16}{11} + \frac{5}{16} \cdot \log_2 \frac{16}{5} = 0.89$; $R = \frac{H}{\sigma} = \frac{0.89}{0.5 \text{ mseg}} = 1\,790$ bits/seg

Para explicar la discrepancia entre este valor de 1 790 bits/seg y el de 1 750 bits/seg hallado anteriormente debemos recordar que la ecuación de entropía se basa en símbolos estadísticamente independientes y mientras esto puede llegar a ser cierto para los pulsos de la figura II.5, no lo es para los sucesivos pulsos binarios, puesto que codificamos en grupos de dos. Para el caso de símbolos dependientes debemos modificar la medida de la información.

e) Un código más eficiente es el siguiente:

Núm. de código	Probabilidad	Código binario
(A) 0	1/2	0
(B) 1	1/4	10
(C) 2	1/8	110
(D) 3	1/8	111

Con este código un mensaje conteniendo $N \gg 1$ símbolos requiere la transmisión de

$$\frac{N}{2} + \frac{2N}{4} + \frac{3N}{8} + \frac{3N}{8} = 1.75 \text{ símbolos}$$

es decir, 1.75 dígitos binarios por símbolo de la fuente. La tasa requerida de señalización es $r = 1\,750$ bits/seg, y así logramos una transmisión a $R = C$, o sea la codificación ha producido una perfecta adaptación entre la fuente y el canal (1 bits/bit).

Hay que notar dos cosas muy interesantes. Primero, que el código es tal que los símbolos 0 y 1 son equiprobables y estadísticamente independientes (verificarlo); segundo, que al símbolo de la fuente con más alta probabilidad se le asigna la palabra del código más corta, y que la longitud crece al ir disminuyendo la probabilidad de los símbolos. Esto es lo que básicamente hizo Morse para su código telegráfico.

Otra cosa interesante es que el código es descifrable sin necesidad de espacios o comas que separen los grupos de códigos. Por ejemplo:

0	10	110	0	111
A	B	C	A	D

Sólo puede significar *ABCAD*. Hay varios métodos para codificar una fuente de salida (Shannon-Fano; Huffman, etc.), que explicaremos más adelante.

II.7. INFORMACIÓN PROMEDIO EN SÍMBOLOS DEPENDIENTES

Cabe resaltar que la mayoría de las fuentes prácticas emiten secuencias de símbolos que son estadísticamente dependientes. En telegrafía, por ejemplo, los mensajes que se van a transmitir consisten en una secuencia de letras, números y caracteres especiales.

Estas secuencias no son completamente aleatorias; en general ellas forman oraciones que tienen una estructura estadística del lenguaje. Por ejemplo, la letra *E* aparece más frecuentemente que la letra *Q*; pero la aparición de *Q* implica que la siguiente letra más probable será la *U*; así, la ocurrencia

de una consonante significa que probablemente la siguiente letra será una vocal. Esta dependencia estadística o estructura reduce la cantidad de información que viene de tal fuente comparada con la cantidad de información que proviene de una fuente que emite los mismos símbolos pero en secuencias independientes.

El problema que ahora encararemos es calcular la tasa de información para fuentes discretas que emiten secuencias dependientes de símbolos o de mensajes. Para el análisis supondremos que la fuente de información discreta emite un símbolo cada T segundos. La fuente entrega símbolos pertenecientes a un alfabeto finito de acuerdo con ciertas probabilidades que dependen en general de los símbolos precedentes, así como del símbolo particular en cuestión. Un sistema físico o modelo estadístico de un sistema que produce tal secuencia de símbolos gobernada por un conjunto de probabilidades, se conoce como proceso estocástico o aleatorio. Entonces nosotros podemos considerar una fuente discreta que sea representada por un proceso aleatorio.

Inversamente, cualquier proceso aleatorio que produzca una secuencia discreta de símbolos elegidos de un conjunto finito puede considerarse como una fuente discreta. Esto incluye los lenguajes escritos como el castellano o el inglés, y también fuentes de información continua que han sido convertidas en discretas por muestreo y cuantificación.

Nosotros podemos hacer un modelo estadístico de la secuencia de símbolos emitidos por la fuente discreta por un proceso estacionario discreto llamado de Markov. Aquí definiremos la entropía y la tasa de información de fuentes de Markov. Supongamos que la fuente puede ser un modelo de un proceso de Markov de estado finito discreto; además, supongamos que el proceso es ergódico, por lo que los promedios en el tiempo pueden ser aplicados. La suposición de ergodicidad implica que el proceso es estacionario, de aquí la probabilidad $P_i(K) = P_i(K+j)$ para cualquier valor de K y j . En otras palabras, la probabilidad de estar en el estado i en el comienzo del intervalo del primer símbolo es la misma, como la probabilidad de estar en el estado i al comienzo del intervalo del segundo símbolo, etc. La probabilidad de ir del estado i al j no depende del tiempo.

Nosotros definimos la entropía de la fuente como un promedio ponderado de la entropía de los símbolos emitidos de cada estado, donde la entropía de estado i indicada por H_i se define como el contenido de información promedio de los símbolos emitidos desde el estado i .

$$H_i = \sum_{j=1}^n P_{ij} \cdot \log_2 (1/p(j/i)) \text{ bits/símbolo}$$

donde $P(j/i)$ es la probabilidad de j habiendo aparecido i . La entropía de la fuente es entonces el promedio de cada estado. Es decir

$$H = \sum_{i=1}^n P_i \cdot H_i = \sum_{i=1}^n P_i \cdot \left[\sum_{j=1}^n P(j/i) \cdot \log_2 1/P(j/i) \right] \text{ bits/símbolo}$$

donde P_i es la probabilidad de la fuente en el estado i . Recordando el teorema de Bayes

$$P(A) \cdot P(B/A) = P(B) \cdot P(A/B) = P(A, B)$$

donde $P(A, B)$ es la probabilidad conjunta de A y B tenemos

$$H = \sum_i \cdot \sum_j P_i \cdot P(j/i) \cdot \log_2 \frac{1}{P(j/i)} = \sum_i \cdot \sum_j P(i, j) \cdot \log_2 \frac{1}{P(j/i)} \quad (\text{II.11})$$

donde P_i es la probabilidad de que la fuente esté en el estado i .

La tasa de información

$$R = r \cdot H \text{ bits/seg} \quad (\text{II.12})$$

donde r es el número de transiciones de estado por segundo o la tasa de símbolos de la fuente.

La entropía H definida de la ecuación (II.11) tiene el mismo significado que H de la ecuación (II.7). En ambos casos nosotros esperamos que la fuente de salida lleve H bits de información por símbolo en mensajes largos.

II.8. FUENTES CONTINUAS

Las fuentes continuas son aquellas cuyos mensajes son funciones que varían continuamente con el tiempo. Tendríamos que definir su entropía pero no lo haremos, pues matemáticamente es pesado, y en cambio le aplicamos en la mayoría de los casos la presentación discreta, puesto que por las limitaciones físicas fundamentales la comunicación es un proceso discreto sin consideración de la fuente.

Ejemplo II.21

Haremos aplicaciones de lo estudiado sobre la estructura del idioma inglés. La información que disponemos en este caso son palabras, es decir, una tira de letras en sucesión que constituyen un texto coherente y que quiero transmitir. Supongamos que utilizo 26 letras y un espacio, por lo tanto 27 símbolos.

Hacemos distintas suposiciones teóricas que nos darán diferentes modelos.

Primer modelo:

Suposiciones: *a*) cada símbolo o letra no depende del anterior; *b*) todos los símbolos o letras son equiprobables, o sea, es simplemente una fuente sin memoria con $m = 27$.

Luego, la entropía tiene en este caso el valor máximo, o sea

$$H = \log_2 27 = 4.75 \text{ bits/letra}$$

Este resultado es importante, ya que me da la carga que tengo que poner en el canal, es decir, mi sistema de transporte debe llevar 4.75 bits por letra.

Segundo modelo:

Suposiciones: *a*) cada símbolo no depende del anterior; *b*) considero a los símbolos con sus probabilidades de aparición.

$$P(\text{espacio}) = 0.1858 \quad P(A) = 0.0642 \quad P(B) = 0.0127 \quad P(z) = 0.0005$$

Calculando tenemos que $H = 4.03$ bits/símbolo

Tercer modelo:

Suposiciones: *a)* cada letra depende solamente del último símbolo transmitido; *b)* probabilidades reales.

En este caso la entropía sigue bajando y vale $H = 3.32$ bits/símbolo.

Cuarto modelo:

Suposiciones: *a)* cada letra depende de las dos anteriores; *b)* probabilidades reales.

La entropía sigue bajando y es $H = 3.10$ bits/símbolo

Quinto modelo:

Suposiciones: *a)* conozco todo el texto anterior; *b)* probabilidades reales.

En este caso $H = 1$ bit/símbolo. Este dato es interesante, ya que si tengo la facilidad en el sistema de transmisión de recordar las letras anteriores y de codificar, la transmisión que se va a enviar, la información que se adiciona al producir una nueva letra, es de 1 bit.

Es decir, que si el sistema es muy eficiente para mandar una letra al canal, debo mandar sólo 1 bit. Este último modelo es el más eficiente, pues el canal debe transportar menos información. Por ejemplo el primer modelo debe estar preparado para llevar 5 bits/letra, y este último modelo sólo debe llevar el canal 1 bit/letra; puede llevar cinco veces más información o ser cinco veces más rápido que el primer modelo.

CUADRO II.1. Frecuencia de las letras en el idioma inglés

Espacio	0.1859	N	0.0574
A	0.0642	O	0.0632
B	0.0127	P	0.0152
C	0.0218	Q	0.0008
D	0.0317	R	0.0484
E	0.1031	S	0.0514
F	0.0208	T	0.0796
G	0.0152	U	0.0008
H	0.0467	V	0.0083
I	0.0575	W	0.0075
J	0.0008	X	0.0013
K	0.0049	Y	0.0164
L	0.0321	Z	0.0005
M	0.0198		

En la práctica se aprovecha el estudio teórico de estos modelos. El primer modelo se utiliza normalmente en teletipos por ser el más económico. Por lo regular se usan cinco dígitos binarios por letra cuando sólo sería necesario en promedio un dígito binario por símbolo si se utilizara el modelo

número 5. El segundo modelo se utiliza en el código Morse de puntos y rayas. Los modelos posteriores (3 y 4) se utilizan en sistemas muy sofisticados de comunicaciones como es el caso de comunicaciones espaciales. En el futuro puede ser que sea necesario el modelo 5. El estudio descrito en este ejercicio lo hizo Shannon en 1951.

Ejemplo II.22

Una fuente binaria de datos tiene $P(0) = 3/8$ y $P(1) = 5/8$, extendiéndose la influencia entre símbolos sobre grupos de dos símbolos sucesivos tal que $P(1/0) = 3/4$ y $P(0/1) = 1/16$. Calcule la entropía condicional H y compare con H_{\max} .

$$P(1/1) = 1 - P(0/1) = 1 - 1/16 = 15/16$$

$$P(0/0) = 1 - P(1/0) = 1 - 3/4 = 1/4$$

$$H = P_0 \left[P(1/0) \cdot \log_2 \frac{1}{P(1/0)} + P(0/0) \cdot \log_2 \frac{1}{P(0/0)} \right] +$$

$$P_1 \left[P(1/1) \cdot \log_2 \frac{1}{P(1/1)} + P(0/1) \cdot \log_2 \frac{1}{P(0/1)} \right] =$$

$$3/8 [3/4 \cdot \log_2 4/3 + 1/4 \cdot \log_2 4] + 5/8 [15/16 \cdot \log_2 16/15 + 1/16 \cdot \log_2 16] = 0.51 \text{ bits}$$

$$H_{\max} = \log_2 2 = 1 \text{ bit}$$



Figura II.6. Influencia de dos símbolos sucesivos

Observe que:

$$P(0, 0) = \frac{3}{8} \cdot \frac{1}{4} = 0.09375 \quad P(0, 1) = \frac{5}{8} \cdot \frac{1}{16} = 0.039$$

$$P(1, 0) = \frac{3}{8} \cdot \frac{3}{4} = 0.281 \quad P(1, 1) = \frac{5}{8} \cdot \frac{15}{16} = 0.581$$

cuya suma de las cuatro probabilidades aproximadamente es 1, y que el criterio aplicado para calcular H es realizar la sumatoria de las informaciones que aporta cada símbolo, afectados por su probabilidad de ocurrencia.

II.9. CANALES DE COMUNICACIÓN

Como ya hemos visto, podemos dividir un sistema de comunicación práctico en un transmisor, un canal físico o un medio de transmisión y un receptor. El transmisor consiste en un codificador y un modulador, mientras el receptor consiste en un demodulador y un decodificador.

El término *canal de comunicación* tiene diferentes significados y caracterizaciones dependiendo de sus puntos terminales y funcionalidad. Entre los puntos *c* y *g* en la figura II.7 tenemos un canal discreto, a menudo conocido como un canal de codificación, que acepta una secuencia de símbolos en su entrada y produce una secuencia de símbolos en su salida. Este canal es caracterizado completamente por un juego de probabilidades de transición $P(j/i)$, donde $P(j/i)$ es la probabilidad de que la salida del canal sea el símbolo *j* del alfabeto cuando a la entrada del canal se tenga el símbolo *i*. Estas probabilidades dependerán de los parámetros del modulador, medio de transmisión, ruido y demodulador.

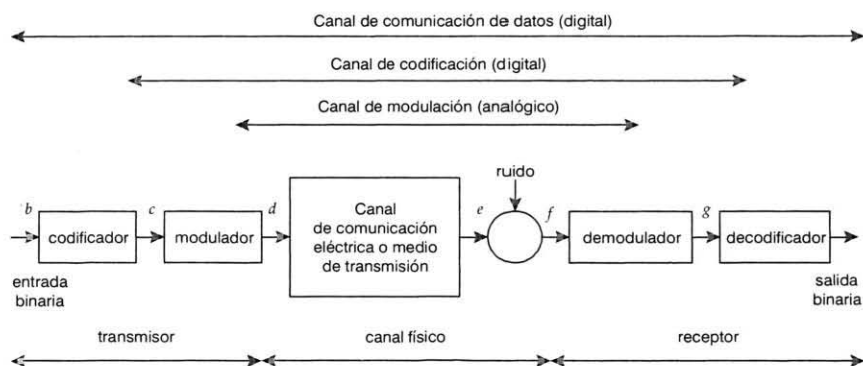


Figura II.7. Caracterización de un canal de comunicación binario

Sin embargo, esta dependencia es transparente a un diseñador del sistema que está enfrascado con el diseño del codificador y decodificador digital.

El canal de comunicación entre los puntos *d* y *f* provee la conexión eléctrica entre el transmisor y el receptor. La entrada y la salida son formas de onda eléctrica analógicas. Esta porción del canal es llamada canal de modulación o canal continuo. Ejemplos de canales de comunicación eléctricos analógicos son los de banda de voz, telefonía, sistemas de radio de alta frecuencia y sistemas troposféricos. Estos canales están sujetos a diversas variedades de contaminación. Algunas debido a variaciones de la respuesta de amplitud y frecuencia del canal dentro de su banda de paso. Otras contaminaciones o degradaciones del canal son debidas a variaciones de las características del canal con el tiempo y no linealidades en el canal. Esto da como resultado que el canal modifica la señal de entrada de una manera determinística (aunque no necesariamente conocida). Además, el canal puede contaminar la señal estadísticamente debido a varios tipos de ruidos aditivos y multiplicativos y desvanecimientos (cambios de atenuación aleatorios dentro del medio de transmisión).

Todas estas contaminaciones introducen errores en la transmisión de datos y limitan la tasa máxima a la que pueden transferirse los datos sobre el canal.

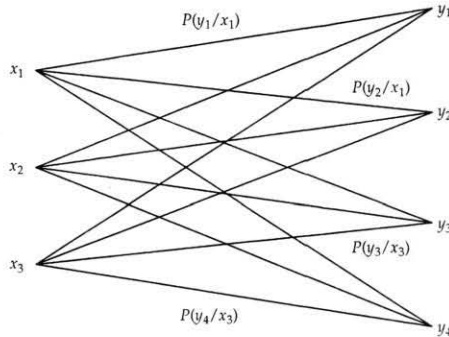


Figura II.8. Diagrama de un canal con tres entradas y cuatro salidas

La capacidad de un canal es uno de los parámetros más importantes de un sistema de comunicación de datos desde que representa la tasa máxima a la que los datos pueden transferirse entre dos puntos en el sistema, con una probabilidad de error arbitrariamente pequeña.

2.10. REPRESENTACIONES DEL CANAL (DISCRETO)

Imaginemos un canal sin memoria, es decir, que la salida del canal en un tiempo dado es una función de su entrada en ese tiempo, y no es una función de las entradas previas del canal. Los canales discretos sin memoria los podemos especificar por un conjunto de probabilidades condicionales que relacionan la probabilidad de cada estado de salida a las probabilidades de entrada (véase Ziemer y Tranter).

En la figura II.8 se muestra un diagrama de un canal con tres entradas y cuatro salidas. Cada trayecto de entrada a salida se indica con una probabilidad condicional $P(y_j/x_i)$, es decir la probabilidad de obtener a la salida y_j si la entrada es x_i . Esto se llama probabilidad de transición del canal.

Especificamos el canal por una matriz de probabilidades de transición $P[y/x]$, donde para el caso de la figura II.8 es:

$$[P(y/x)] = \begin{bmatrix} p(y_1/x_1) & p(y_2/x_1) & p(y_3/x_1) & p(y_4/x_1) \\ p(y_1/x_2) & p(y_2/x_2) & p(y_3/x_2) & p(y_4/x_2) \\ p(y_1/x_3) & p(y_2/x_3) & p(y_3/x_3) & p(y_4/x_3) \end{bmatrix} \quad (\text{II.13})$$

Notemos que la suma de cada renglón de la matriz es la unidad.

La matriz del canal es útil para encontrar las probabilidades de salida dadas las probabilidades de entrada. Por ejemplo, si las probabilidades de entrada se representan por la matriz:

$$[P(x)] = [p(x_1) \ p(x_2) \ p(x_3)] \quad (\text{II.14})$$

entonces

$$[P(y)] = [p(y_1) p(y_2) p(y_3) p(y_4)] \quad (\text{II.15})$$

luego, teniendo en cuenta la ecuación (II.13).

$$P[(y)] = P[(x)] \cdot P[(y/x)] \quad (\text{II.16})$$

$$[P(y)] = [p(x_1)p(x_2)p(x_3)] \cdot \begin{bmatrix} p(y_1/x_1) & p(y_2/x_1) & p(y_3/x_1) & p(y_4/x_1) \\ p(y_1/x_2) & p(y_2/x_2) & p(y_3/x_2) & p(y_4/x_2) \\ p(y_1/x_3) & p(y_2/x_3) & p(y_3/x_3) & p(y_4/x_3) \end{bmatrix}$$

Si $P(x)$ se escribe como una matriz en diagonal, es decir,

$$\begin{bmatrix} p(x_1) & 0 & 0 \\ 0 & p(x_2) & 0 \\ 0 & 0 & p(x_3) \end{bmatrix}$$

y recordando que el producto de una matriz A con m filas y S columnas por una matriz B de S filas y n columnas es una matriz C de m filas y n columnas, y observando por Bayes que

$$[p(x_i) \cdot p(y_j/x_i) = p(x_i, y_j)]$$

hace que $[P(y)]$ se conozca con el nombre de matriz de probabilidad conjunta y el término $p(x_i, y_j)$ es la probabilidad conjunta de transmitir x_i y recibir y_j .

Ejemplo II.23

Consideremos el canal binario de la figura II.9. La matriz de probabilidades de transición es

$$p[y/x] = \begin{bmatrix} p(y_1/x_1) & p(y_2/x_1) \\ p(y_1/x_2) & p(y_2/x_2) \end{bmatrix} = \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}$$

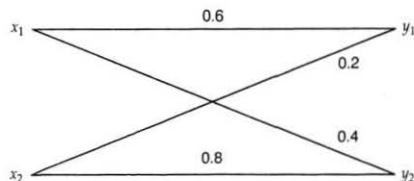


Figura II.9. Canal binario

Si $P(x_1) = 0.5$, las probabilidades de salida son:

$$P(y) = [P(x)] \cdot [P(y/x)]$$

$$P(y) = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix} \cdot \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}$$

$$P(y) = [0.4 \ 0.6] = [p(y_1) \ p(y_2)]$$

y la matriz de probabilidad conjunta es

$$P[x, y] = [P(x) \text{ en diagonal}] \cdot [P(y/x)]$$

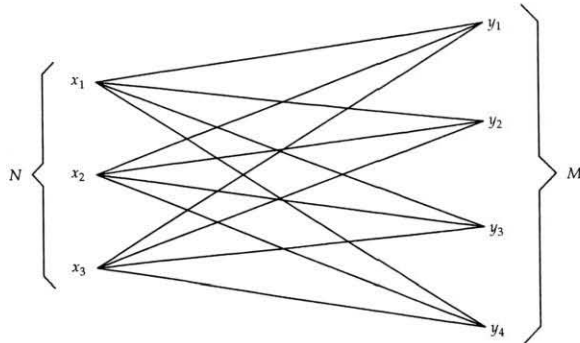
$$P[x, y] = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \cdot \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix} \downarrow$$

donde las flechas indican cómo deben multiplicarse las matrices

$$P[x, y] = \begin{bmatrix} 0.3 & 0.2 \\ 0.1 & 0.4 \end{bmatrix} = \begin{bmatrix} 0.6 \cdot 0.5 & 0.5 \cdot 0.4 \\ 0.5 \cdot 0.2 & 0.5 \cdot 0.8 \end{bmatrix} = \begin{bmatrix} p(x_1, y_1) & p(x_1, y_2) \\ p(x_2, y_1) & p(x_2, y_2) \end{bmatrix}$$

Ejemplo II.24

Para el canal de la figura, verifique que si las entradas son equiprobables las salidas también lo son.



$$[P(y/x)] = \begin{bmatrix} p(y_1/x_1) & p(y_2/x_1) & p(y_3/x_1) & p(y_4/x_1) \\ p(y_1/x_2) & p(y_2/x_2) & p(y_3/x_2) & p(y_4/x_2) \\ p(y_1/x_3) & p(y_2/x_3) & p(y_3/x_3) & p(y_4/x_3) \end{bmatrix}$$

$$P[y] = P[x] \cdot P[y/x] = [p(x_1) \ p(x_2) \ p(x_3)] \cdot [P(y/x)]$$



Tomamos una salida $y_j = y_2$

$$P[y_2] = [p(x_1)p(x_2)p(x_3)] \cdot \begin{bmatrix} p(y_2/x_1) \\ p(y_2/x_2) \\ p(y_2/x_3) \end{bmatrix} = \sum_{i=1}^N p(y_2/x_i)p(x_i)$$

Si las salidas son equiprobables

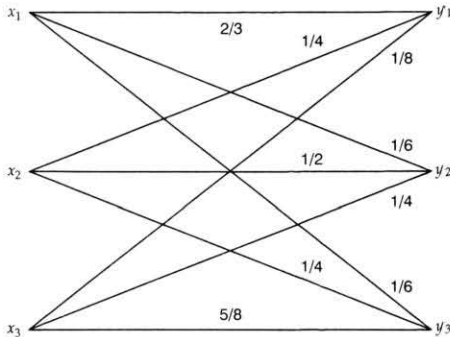
$$P[y_j] = \frac{1}{N} \cdot \left[\sum_{i=1}^N p(y_j/x_i) \right]$$

$$P[y] = [0.33 \quad 0.33 \quad 0.33 \quad 0.33] \cdot \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix} =$$

$$= [0.25 \quad 0.25 \quad 0.25 \quad 0.25] = [y_1 \quad y_2 \quad y_3 \quad y_4]$$

Ejemplo II.25

Tenemos un canal con su matriz de probabilidades de transición; determine $P(y)$ para entradas equiprobables y la matriz de probabilidad conjunta.



$$P[y/x] = \begin{bmatrix} 2/3 & 1/6 & 1/6 \\ 1/4 & 1/2 & 1/4 \\ 1/8 & 1/4 & 5/8 \end{bmatrix}$$

Para entradas equiprobables:

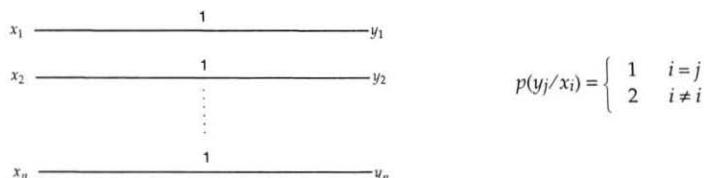
$$P[y] = [1/3 \quad 1/3 \quad 1/3] \cdot \begin{bmatrix} 2/3 & 1/6 & 1/6 \\ 1/4 & 1/2 & 1/4 \\ 1/8 & 1/4 & 5/8 \end{bmatrix} = [25/72 \quad 22/72 \quad 25/72]$$

La matriz de probabilidad conjunta:

$$P[x, y] = \begin{bmatrix} 1/3 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/3 \end{bmatrix} \cdot \begin{bmatrix} 2/3 & 1/6 & 1/6 \\ 1/4 & 1/2 & 1/4 \\ 1/8 & 1/4 & 5/8 \end{bmatrix} = \begin{bmatrix} 2/9 & 1/18 & 1/18 \\ 1/12 & 1/6 & 1/12 \\ 1/24 & 1/12 & 5/24 \end{bmatrix}$$

Ejemplo II.26

Descubra la matriz de probabilidad de transición para el canal sin ruido de la figura.



$$P[x/y] = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

$$p(x_i, y_j) = P(x_i) \cdot P(y_j/x_i) = \begin{cases} P(x_i) & i=j \\ 0 & i \neq j \end{cases}$$

$$P[x/y] = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

II.11. ENTROPÍA CONDICIONAL Y CONJUNTA

Usando las probabilidades de entrada y salida $p(x_i)$ y $p(y_j)$, las probabilidades de transición $P(y_j/x_i)$ y las probabilidades conjuntas $P(x_i, y_j)$ podemos definir diferentes funciones de entropía para un canal con n entrada y m salidas:

$$H(x) = - \sum_{i=1}^n p(x_i) \cdot \log_2 p(x_i) \quad (\text{II.17})$$

que es la incertidumbre promedio de la fuente. La incertidumbre promedio del símbolo recibido es

$$H(y) = - \sum_{j=1}^n p(y_j) \cdot \log_2 p(y_j) \quad (\text{II.18})$$

Observando la fórmula (II.11) tenemos la incertidumbre promedio del símbolo recibido para cuando sea transmitido x

$$H(y/x) = - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \cdot \log_2 p(y_j/x_i) \quad (\text{II.19})$$

y la incertidumbre promedio del símbolo transmitido después de haber recibido un símbolo

$$H(x/y) = - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \cdot \log_2 p(x_i/y_j) \quad (\text{II.20})$$

a veces llamada **Equivocación**.

La entropía conjunta $H(x, y)$ es la incertidumbre promedio del sistema de comunicación

$$H(x, y) = - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \cdot \log_2 p(x_i, y_j) \quad (\text{II.21})$$

Por el teorema de Bayes podemos escribir

$$\begin{aligned} H(x, y) &= - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \cdot [\log_2 p(x_i/y_j) + \log_2 p(y_j)] \\ &= - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \cdot \log_2 p(x_i/y_j) \\ &\quad - \sum_{j=1}^m p(y_j) \cdot \log_2 p(y_j) = H(x/y) + H(y) \end{aligned} \quad (\text{II.22})$$

De manera similar podemos demostrar que

$$H(x, y) = H(y/x) + H(x) \quad (\text{II.23})$$

II.12. CAPACIDAD DEL CANAL

Consideremos un observador en la salida de un canal. La incertidumbre promedio del observador respecto a la entrada del canal tendrá un cierto valor antes de la recepción de la salida y su incertidumbre promedio de la entrada disminuye cuando se recibe la salida. Dicho de otra manera

$$H(x/y) \leq H(x)$$

La disminución es la incertidumbre promedio del observador de la señal transmitida, cuando la salida se recibe es una medida de la información media transmitida. Esto se conoce como información mutua $I(x, y)$. Entonces

$$I(x, y) = H(x) - H(x/y) \quad (\text{II.24})$$

pero de (II.22) y de (II.23)

$$H(x, y) = H(x/y) + H(y) = H(y/x) + H(x)$$

luego

$$I(x, y) = H(y) - H(y/x) = H(x) - H(x/y) \quad (\text{II.25})$$

Debemos observar que la información mutua es una función de las probabilidades de la fuente, así como de las probabilidades de transición del canal.

La capacidad del canal C se define como el máximo valor de información promedio por símbolo que puede transmitirse por el canal. Por lo tanto

$$C = \text{máxima } [I(x, y)] \quad (\text{II.26})$$

La maximización es respecto a las probabilidades de la fuente desde que las probabilidades de transición están fijadas por el canal. Sin embargo, la capacidad del canal es solamente una función de las probabilidades de transición del canal, a partir de que el proceso de maximización elimina la dependencia de las probabilidades de la fuente (véase Abramson).

Ejemplo II.27

Encuentre la capacidad del canal discreto sin ruido de la figura II.10

Sabemos de la fórmula (II.25) que $I(x, y) = H(x) - H(x/y)$, y de la (II.20)

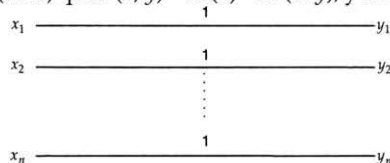


Figura II.10. Canal discreto sin ruido

$$H(x/y) = - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \cdot \log_2 p(x_i/y_j)$$

Para el canal sin ruido todos los $P(x_i, y_j)$ y $P(x_i/y_j)$ son cero excepto para $i = j$ que vale la unidad. Entonces $H(x/y) = 0$ e

$$I(x, y) = H(x)$$

Sabemos que la entropía de una fuente es máxima si todos los símbolos de la fuente son igualmente probables. Por lo tanto

$$C = \sum_{i=1}^n \frac{1}{n} \log_2 n = \log_2 n \text{ (bits/símbolo)} \quad (\text{II.27})$$

Si $n = 2$, entonces $C = 1$ bit/símbolo.

Si la velocidad de señalización de la fuente de entrada es r (símbolos/seg), la capacidad del canal binario sin ruido en bit/seg es

$$C = r \log_2 n \quad (\text{II.28})$$

Ejemplo II.28

Llegaremos a la expresión (II.28) por otro camino. Hay que destacar que la tasa de información promedio da la cantidad de información producida por una fuente en un tiempo dado, mientras que la capacidad del canal C en bits por segundo es una medida de la cantidad de información que un canal puede transferir por unidad de tiempo.

Hay un teorema, a veces conocido como el primer teorema de Shannon, que dice:

“Dado un canal de capacidad C y una fuente con entropía R , si $R \leq C$ existe una técnica de codificación tal que la salida de la fuente puede ser transmitida sobre el canal con una frecuencia pequeña de errores a pesar de la presencia del ruido. Si $R > C$ no es posible transmitir sin errores”.

La demostración de este teorema se puede ver en Shannon. Si $M(T)$ es el número de mensajes diferentes de longitud T que pueden identificarse en el receptor con una baja probabilidad de error, entonces:

$$C = \lim_{T \rightarrow \infty} \frac{1}{T} \cdot \log_2 M(T) \text{ bits/seg} \quad (\text{II.29})$$

Recordemos que un canal discreto es aquel que transmite información tomando un número discreto de estados eléctricos, voltajes, etc., siendo n el número de posibles estados y r la tasa de señalización en estado por unidad de tiempo. Si la relación señal a ruido es grande la probabilidad de error es pequeña y el canal puede considerarse como sin ruido (cualquier secuencia de símbolos se interpreta correctamente).

$$\text{Entonces: } C = \lim_{T \rightarrow \infty} \frac{1}{T} \cdot \log_2 n^T = r \log_2 n \text{ bits/seg}$$

lo cual es igual a lo obtenido en (II.28).

Cuando el ruido del canal no puede dejar de considerarse, la capacidad es menor a causa de los errores. Esto podrá observarse en el ejercicio siguiente.

Ejemplo II.29

Encuentre la capacidad del canal binario simétrico (CBS) de la figura

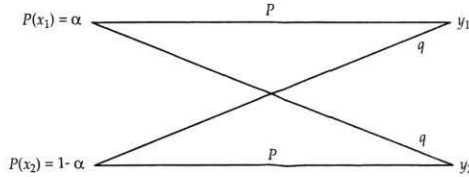


Figura II.11. Canal binario simétrico

Determinaremos la capacidad maximizando

$$I(x, y) = H(y) - H(y/x) \text{ donde}$$

$$H(y/x) = - \sum_{i=1}^2 \sum_{j=1}^2 p(x_i, y_j) \cdot \log_2 p(y_j/x_i)$$

$$= -p(x_1, y_1) \log_2 p(y_1/x_1) - p(x_1, y_2) \log_2 p(y_2/x_1)$$

$$- p(x_2, y_1) \log_2 p(y_1/x_2) - p(x_2, y_2) \log_2 p(y_2/x_2)$$

$$H(y/x) = -p(y_1/x_1) p(x_1) \log_2 p(y_1/x_1) - p(y_2/x_1) p(x_1) \log_2 p(y_2/x_1)$$

$$- p(y_1/x_2) p(x_2) \log_2 p(y_1/x_2) - p(y_2/x_2) p(x_2) \log_2 p(y_2/x_2)$$

$$= -[p \cdot \alpha \cdot \log_2 p] - [q \cdot \alpha \cdot \log_2 q] - [q \cdot (1 - \alpha) \cdot \log_2 q]$$

$$- [p(1 - \alpha) \log_2 p] = -p \log_2 p - q \log_2 q$$

$$I(x, y) = H(y) - H(y/x) = H(y) + p \cdot \log_2 p + q \cdot \log_2 q$$

Esto es máximo para $H(y)$ máxima o sea $H(y) \approx 1$. Entonces la capacidad del canal binario simétrico es

$$C = 1 + p \log_2 p + q \log_2 q \quad (\text{II.30})$$

que se encuentra graficado en la figura II.12.

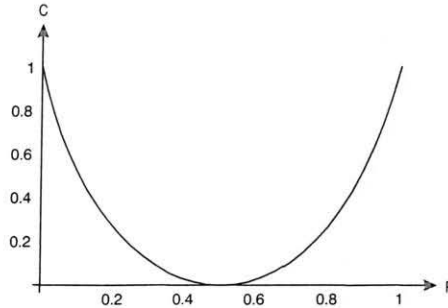


Figura II.12. Capacidad de un canal simétrico

Si $p = 0$ o 1 , la salida del canal es completamente determinada por el canal de entrada, y la capacidad es de 1 bit/símbolo. Si $P = 1/2$ un símbolo de entrada produce cualquier salida con la misma probabilidad y la capacidad es cero.

Ejemplo II.30

Determine la probabilidad de error para un CBS (canal binario simétrico).

La probabilidad de error es

$$P_E = \sum_{i=1}^2 p(e/x_i) \cdot p(x_i) \quad (\text{II.31})$$

donde $p(e/x_i)$ es la probabilidad de error dada la entrada x_i . Entonces

$$P_E = q \cdot p(x_1) + q \cdot p(x_2) = q \quad (\text{II.32})$$

o sea, igual a la probabilidad de error adicional

$$P(y_j/x_i) \text{ para } i \neq j$$

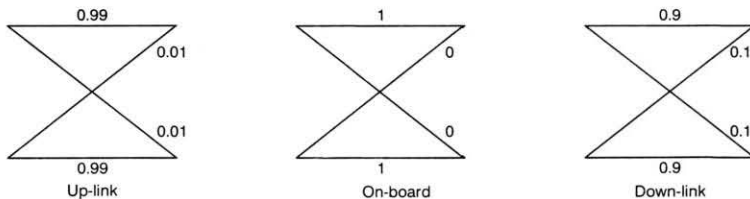
Más adelante demostraremos que la P_E disminuye con la energía de los símbolos recibidos. Puesto que la energía de un símbolo es la potencia recibida multiplicada por el periodo del símbolo, se desprende que si la potencia del transmisor es fija, la probabilidad de error puede reducirse disminuyendo la tasa de símbolos de la fuente.

Ejemplo II.31

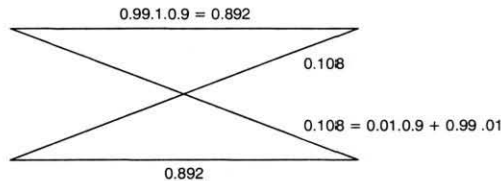
Un sistema de satélite consta de tres canales. El canal estación terrena-satélite (*up-link*) con $P_{\text{error}} = 0.01$; el canal de procesamiento de la señal a bordo del satélite (*on-board*) con $P_{\text{error}} = 0$ y el canal satélite-estación-terrena (*down-link*) con $P_{\text{error}} = 0.1$. La probabilidad del *down-link*, es mayor que la del *up-link* debido a la limitación de potencia en el satélite. Calcule la capacidad total del sistema.

Respuesta:

El sistema se puede considerar como 3 CBS.



Los que dan un CBS equivalente



Las capacidades de cada uno de los canales y la capacidad total son

$$C_1 = 1 + 0.01 \log_2 0.01 + 0.99 \log_2 0.99 = 0.9192 \text{ bits/símbolo}$$

$$C_2 = 1 \text{ bit/símbolo}$$

$$C_3 = 0.530 \text{ bits/símbolo}$$

y la capacidad total

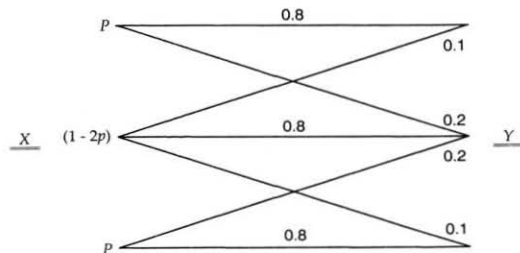
$$C = 1 + 0.89 \log_2 0.892 + 0.108 \log_2 0.108 = 0.561 \text{ bits/símbolo}$$

En la práctica, las probabilidades son menores. Se tomaron valores mayores por facilidades de cálculo. Otra manera de determinar la capacidad total es:

$$\begin{aligned}
 P[y/x] &= \begin{bmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix} = \\
 &= \begin{bmatrix} 0.99 \cdot 1 + 0.01 \cdot 0 & 0.99 \cdot 0 + 0.01 \cdot 1 \\ 0.01 \cdot 1 + 0.99 \cdot 0 & 0.01 \cdot 0 + 0.99 \cdot 1 \end{bmatrix} \cdot \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix} = \\
 &= \begin{bmatrix} 0.99 \cdot 0.9 + 0.01 \cdot 0.1 & 0.01 \cdot 0.9 + 0.99 \cdot 0.1 \\ 0.01 \cdot 0.9 + 0.99 \cdot 0.1 & 0.01 \cdot 0.1 + 0.99 \cdot 0.9 \end{bmatrix} = \begin{bmatrix} 0.892 & 0.108 \\ 0.108 & 0.892 \end{bmatrix}
 \end{aligned}$$

Ejemplo II.32

Encuentre la capacidad del canal discreto de la figura



$$H(y) = - \sum_{j=1}^m p(y_j) \cdot \log_2 p(y_j)$$

$$H(y) = - [0.8p + 0.1 - 0.2p] \log_2 [0.8p + 0.1 - 0.2p]$$

$$- [0.2p + 0.8 - 1.6p + 0.2p] \log_2 (0.8 - 1.2p)$$

$$- [0.1 - 0.2p + 0.8p] \log_2 [0.1 + 0.6p]$$

$$= - [0.6p + 0.1] \log_2 [0.6p + 0.1] - [0.8 - 1.2p] \log_2 [0.8 - 1.2p]$$

$$- [0.6p + 0.1] \log_2 [0.6p + 0.1]$$

$$= - [1.2p + 0.2] \log_2 [0.6p + 0.1] - [0.8 - 1.2p] \log_2 [0.8 - 1.2p]$$

$$H(y/x) = - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \cdot \log_2 p(y_j/x_i)$$

$$H(y/x) = - p(x_1, y_1) \log_2 p(y_1/x_1) - p(x_1, y_2) \log_2 p(y_2/x_1)$$

$$- p(x_1, y_3) \log_2 p(y_3/x_1) - p(x_2, y_1) \log_2 p(y_1/x_2)$$

$$\begin{aligned}
 & -p(x_2, y_2) \log_2 p(y_2/x_2) - p(x_2, y_3) \log_2 p(y_3/x_2) \\
 & -p(x_3, y_1) \log_2 p(y_1/x_3) - p(x_3, y_2) \log_2 p(y_2/x_3) \\
 & -p(x_3, y_3) \log_2 p(y_3/x_3) \\
 = & -p(y_1/x_1) p(x_1) \log_2 p(y_1/x_1) - p(y_2/x_1) p(x_1) \log_2 p(y_2/x_1) \\
 & -p(y_3/x_1) p(x_1) \log_2 p(y_3/x_1) - p(y_1/x_2) p(x_2) \log_2 p(y_1/x_2) \\
 & -p(y_2/x_2) p(x_2) \log_2 p(y_2/x_2) - p(y_3/x_2) p(x_2) \log_2 p(y_3/x_2) \\
 & -p(y_1/x_3) p(x_3) \log_2 p(y_1/x_3) - p(y_2/x_3) p(x_3) \log_2 p(y_2/x_3) \\
 & -p(y_3/x_3) p(x_3) \log_2 p(y_3/x_3)
 \end{aligned}$$

Luego

$$y_1/x_1 = 0.8p$$

$$y_1/x_2 = 0.1 - 0.2p$$

$$y_1/x_3 = 0$$

$$y_2/x_1 = 0.2p$$

$$y_2/x_2 = 0.8 - 1.6p$$

$$y_2/x_3 = 0.2p$$

$$y_3/x_1 = 0$$

$$y_3/x_2 = 0.1 - 0.2p$$

$$y_3/x_3 = 0.8p$$

$$p(x_1) = p$$

$$p(x_2) = 1 - 2p$$

$$p(x_3) = p$$

Reemplazando valores:

$$H(y/x) = -0.4p + 0.4$$

Luego

$$\begin{aligned}
 I(x, y) &= H(y) - H(y/x) = [0.2 + 1.2p] \log_2 [0.1 + 0.6p] \\
 &\quad - [0.8 - 1.2p] \log_2 [0.8 - 1.2p] - 0.4p + 0.4
 \end{aligned}$$

Sabemos que $C = \max I(X, Y)$

Hallamos el valor de p que da el máximo $I(X, Y)$

$$\frac{dI(x, y)}{dp} = 0 \Rightarrow p = 0.43$$

Sustituyendo el valor de $p = 0.43$ en $I(X, Y)$ tenemos

$$C = 0.84 \text{ bits/símbolo}$$

Si la tasa de señalización es 10 000 símbolos/seg, la capacidad en bits/seg es de

$$0.84 \frac{\text{bits}}{\text{símbolo}} \times 10\,000 \frac{\text{símbolos}}{\text{seg}} = 8\,400 \text{ bits/seg}$$

II.13. EXTENSIÓN DE UNA FUENTE

Tenemos una fuente que emite símbolos de un alfabeto $S = S_1; S_2; S_3$ con $p(S_1) = 1/2$; $P(S_2) = P(S_3) = 1/4$, o sea:

Símbolo	Probabilidad
S1	1/2
S2	1/4
S3	1/4

La entropía de esta fuente es:

$$H(s) = \frac{1}{2} \log_2 2 + 2 \cdot \frac{1}{4} \log_2 4 = 1.5 \text{ bits/símbolo}$$

Haremos la segunda extensión de la fuente $S'' = S^2$, la que tendrá $q'' = 3^2 = 9$ símbolos, o sea:

Símbolos de S^2	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7	σ_8	σ_9
Símbolo de la segunda extensión de la fuente	$S_1 S_1$	$S_1 S_2$	$S_1 S_3$	$S_2 S_1$	$S_2 S_2$	$S_2 S_3$	$S_3 S_1$	$S_3 S_2$	$S_3 S_3$
Probabilidad de $P(\sigma_i)$	1/4	1/8	1/8	1/8	1/16	1/16	1/8	1/16	1/16

En uno de los ejercicios al final de este capítulo se demuestra que la entropía de la extensión misma de una fuente es n por la entropía de la fuente. Es decir:

$$H(S^n) = n \cdot H(S) \quad (\text{II.33})$$

En nuestro caso la segunda extensión de la fuente tiene una entropía

$$H(S^2) = 3 \text{ bits/símbolo} = \frac{1}{4} \cdot \log_2 4 + 4 \cdot \frac{1}{8} \cdot \log_2 8 + 4 \cdot \frac{1}{16} \log_2 16$$

$$= \sum_{i=1}^9 P(\sigma_i) \cdot \log_2 \frac{1}{P(\sigma_i)} = 2 \cdot 1.5 \text{ bits/símbolo}$$

Ejemplo II.33

Consideremos una fuente discreta con dos posibles símbolos de salida A y B con probabilidades $P(A) = 0.8$ y $P(B) = 0.2$. Supongamos que la tasa de símbolos de la fuente es 1 300 símbolos/seg y que el canal puede transmitir 1 000 símbolos/seg.

Si tomamos la codificación del cuadro A vemos que la transmisión es imposible, pues la tasa de salida del codificador es 1 300 símbolos/seg y sólo 1 000 símbolos/seg pueden ser transmitidos por medio del canal.

CUADRO A

Símbolo de la fuente	Probabilidad	Palabra del código	(Probabilidad) \times (núm. de símbolos del código)
A	0.8	0	0.8
B	0.2	1	0.2
Longitud de palabra promedio = 1			

Pero como la capacidad del canal es $C = 1\,000$ bits/seg; $H = 0.72$ bits/símbolo, y para la codificación del cuadro A; $R = 1\,300 \cdot 0.72 = 936$ bit/seg, vemos que entonces es posible encontrar una nueva codificación que haga factible la transmisión por el canal.

Intentaremos codificar cada par de salidas de la fuente y asignaremos la palabra más corta del código a la secuencia más probable. Este esquema se ve en el cuadro B y se conoce como la extensión de segundo orden de la fuente. Con este esquema hay $1.56/2 = 0.78$ símbolos del codificador por símbolo de salida de la fuente. La tasa de símbolos de salida del codificador es

$$0.78 \cdot 1\,300 = 1\,014 \text{ símbolos/seg}$$

que es mayor que la que el canal puede transmitir.

CUADRO B

Símbolo de la fuente	Probabilidad	Palabra del código	(Probabilidad) \times (núm. de símbolos del código)
AA	0.64	0	0.64
AB	0.16	10	0.32
BA	0.16	110	0.48
BB	0.04	111	0.12
Longitud de palabra promedio = $1.56 = 0.64 + 0.32 + 0.48 + 0.12$			

Intentamos con la tercera extensión de la fuente.

CUADRO C

<i>Símbolo de la fuente</i>	<i>Probabilidad</i>	<i>Palabra del código</i>	<i>(Probabilidad) x (número de símbolos del código)</i>
AAA	0.512	0	0.512
AAB	0.128	100	0.384
ABA	0.128	101	0.384
BAA	0.128	110	0.384
ABB	0.032	11100	0.160
BBA	0.032	11101	0.160
BAB	0.032	11110	0.160
BBB	0.008	11111	0.040
Longitud de palabra promedio = 2.184			

Aquí hay $\frac{2.184}{3} \cdot 1\,300 = 946$ símbolos/seg, que ahora el canal binario puede aceptar.

En el cuadro D aparece un resumen de las codificaciones que hemos visto, para una determinada salida de la fuente.

CUADRO D

Fuente	A	B	A	A	A	A	B	A	A	A	A	B	A	B	A	A	A	A	A	A	A	B	A	B	A	A	A	A
Código	0	1	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0
Código	1	0				0	11	0			0	11	0	11	0		0		0		0	1	0	1	0		0	0
Código		10	1			0		11	0			0	11	11	0			0			0	10	0		10	1		0

Las entropías a la salida del codificador son:

$$H_1 = \frac{936 \text{ bits/seg}}{1\,300 \text{ símbolos/seg}} = 0.720 \text{ bit/símbolo}$$

$$H_2 = \frac{936 \text{ bits/seg}}{1\,014 \text{ símbolos/seg}} = 0.923 \text{ bit/símbolo}$$

$$H_3 = \frac{936 \text{ bits/seg}}{946 \text{ símbolos/seg}} = 0.989 \text{ bit/símbolo}$$

recordemos que 936 bits/seg es la tasa de información de la fuente.

Observará que en el último caso la entropía aumenta, puesto que las probabilidades para un 1 o un 0 son casi iguales. El objetivo del codificador es aumentar la entropía de la fuente vista por el canal. Como ejercicio, verifique que en el último caso $P(1) = 0.472$ y $P(0) = 0.528$.

Una medida de la bondad de codificación de la fuente es la eficiencia que se define como la relación entre la longitud promedio mínimo de las palabras del código L_{\min} a la longitud de palabra promedio del código \bar{L} . Así.

$$\text{Eficiencia} = \frac{\bar{L}_{\min}}{\bar{L}} = \frac{\bar{L}_{\min}}{\sum_{i=1}^n p(x_i) \cdot l_i}$$

donde $p(x_i)$ es la probabilidad del i -ésimo símbolo de la fuente y l_i es la longitud de la palabra del código correspondiente al i -ésimo símbolo de la fuente. Puede demostrarse que la longitud de palabra mínima es dada por

$$\bar{L}_{\min} = \frac{H}{\log_2 K}$$

siendo K el número de símbolos del alfabeto de codificación.

Entonces:

$$\text{Eficiencia} = \frac{H}{\bar{L} \cdot \log_2 K} \quad [\text{En este ejercicio } K = 2]$$

La redundancia de un código se define como 1 menos la eficiencia, por lo que es conveniente eliminar redundancia en la codificación para aumentar la eficiencia. Las eficiencias $= H(X^n)$ coinciden con los valores de entropía en nuestro ejercicio. Es decir:

$$\text{Eficiencia 1} = \frac{H(x)}{\bar{L}} = \frac{0.72}{1} = 0.720 \therefore \text{Redundancia} = 28\%$$

$$\text{Eficiencia 2} = \frac{H(x^2)}{\bar{L}} = \frac{2 \cdot 0.72}{1.56} = 0.923 \therefore \text{Redundancia} = 7.7\%$$

$$\text{Eficiencia 3} = \frac{H(x^3)}{\bar{L}} = \frac{3 \cdot 0.72}{2.184} = 0.989 \therefore \text{Redundancia} = 1.1\%$$

Es importante hacer notar que con la extensión de la fuente la entropía aumenta, pero que la tasa de información no se modifica.

$$\text{Primera extensión} \rightarrow (0.72) \cdot 1300 = 936 \text{ bits/seg}$$

$$\text{Segunda extensión} \rightarrow (0.923) \cdot (1014) = 936 \text{ bits/seg}$$

$$\text{Tercera extensión} \rightarrow (0.989) \cdot (946) = 936 \text{ bits/seg}$$

Lo que hemos visto es correcto pues la codificación no puede modificar la tasa de información de la fuente.

II.14. CÓDIGOS INSTANTÁNEOS Y NO INSTANTÁNEOS

Se hace esta clasificación de acuerdo con la posibilidad de decodificar cada palabra sin referencia a los símbolos que siguen. Los códigos no instantáneos requieren referencia de los símbolos que siguen. Hay que notar que ambos códigos son únicamente descifrables. Por ejemplo.

<i>Símbolos de la fuente</i>	<i>Código no instantáneo</i>	<i>Código instantáneo</i>
X_1	0	0
X_2	01	10
X_3	011	110
X_4	0111	1110

Codificación de Shannon-Fano y Huffman

Hay varios métodos para codificar una fuente de salida; aquí consideraremos primero el de Shannon-Fano y luego el de Huffman. En el de Shannon-Fano las salidas de la fuente se dividen en orden de probabilidad decreciente. El conjunto se divide en dos partes (Línea B-B'), tratando de que cada parte tenga la misma probabilidad, o lo más cercana posible. Entonces se le asigna un 0 en la primera columna de la parte superior y un 1 en la parte inferior. Se sigue este proceso con la misma metodología hasta que no es posible obtener más partes o separaciones. Este esquema da un código 100% eficiente si las partes separadas son equiprobables. Para el ejemplo de la figura II.13.

$$\text{Eficiencia} = \frac{H}{L} = \frac{2.75}{2.75} = 1$$

puesto que las partes pueden separarse en forma equiprobable.

<i>Palabras de la fuente</i>	<i>Probabilidad</i>	<i>Palabras del código</i>	<i>(Longitud) x (probabilidad)</i>
1	0.25	00	$2(0.25) = 0.50$
2	0.25	01	$2(0.25) = 0.50$
B..... B'			
3	0.125	100	$3 \cdot (0.125) = 0.375$
4	0.125	101	$3 \cdot (0.125) = 0.375$
5	0.0625	1 100	$4 \cdot (0.0625) = 0.25$
6	0.0625	1 101	$4 \cdot (0.0625) = 0.25$
7	0.0625	1 110	$4 \cdot (0.0625) = 0.25$
8	0.0625	1 111	$4 \cdot (0.0625) = 0.25$

Longitud media de palabra = 2.75

Figura II.13. Codificación de Shannon-Fano

El procedimiento sugerido por Huffman da un código con la longitud de palabra más corta. Este código se conoce como óptimo (véase Abramson). Supongamos que tenemos una fuente con seis símbolos y sus respectivas probabilidades, la queremos codificar con un alfabeto binario por el método de Huffman.

- 1) Ordenamos los símbolos en orden decreciente de probabilidad.
- 2) Unimos los dos últimos símbolos convirtiéndolos en un solo símbolo cuya probabilidad es la suma de ambos.

Volvemos a ordenar los símbolos en orden decreciente de probabilidad.

- 3) Repetimos la unificación hasta llegar a una fuente de dos símbolos.

- 4) Codificamos la última fuente en 0 y 1.

Símbolos	Probabilidad	Fuente reducida			
m_1	0.4	0.4	0.4	0.4	0.6 \Rightarrow 0
m_2	0.3	0.3	0.3	0.3	
m_3	0.1	0.1	0.2	0.3	0.4 \Rightarrow 1
m_4	0.1	0.1	0.1		
m_5	0.06				
m_6	0.04				

El último paso es:

- 5) Codificamos las fuentes o etapas precedentes (se procede desde la fuente de dos símbolos hacia atrás), agregando un 0 o un 1 al símbolo desdoblado, o sea:

Símbolos	Probabilidad	Código	Fuente reducida			
m_1	0.4	1	0.4	1	0.4	1 \Rightarrow 0.6 0
m_2	0.3	00	0.3	00	0.3	00 \Rightarrow 0.4 1
m_3	0.1	0100	0.1	011	0.2	010 \Rightarrow 0.3 01
m_4	0.1	0101	0.1	0100	0.1	011
m_5	0.06	0110	0.1	0101		
m_6	0.04	0111				

La codificación es (código instantáneo):

m_1	1	m_4	0101
m_2	00	m_5	0110
m_3	0100	m_6	0111

II.15. CANALES CONTINUOS

El canal de comunicación entre los puntos d y f de la figura II.7 es analógico o continuo en su naturaleza. En esa parte del canal, las señales de entrada son funciones continuas del tiempo, y la función del canal es producir en su salida la onda eléctrica que se presenta en su entrada. Un canal real lo logra sólo aproximadamente: primero, el canal modifica la forma de onda de una manera determinada, y este efecto puede ponerse en un modelo tratando al canal como un sistema lineal. El canal también modifica aleatoriamente la forma de onda de entrada debido al ruido multiplicativo y aditivo. Nosotros trataremos con este último ya que ocurre más a menudo que el multiplicativo.

El ruido aditivo puede ser gaussiano o impulsivo en su naturaleza. Cuando es gaussiano incluye el ruido térmico y el ruido impacto (*shot noise*) del equipamiento y la radiación captada por la antena receptora. De acuerdo con el teorema del límite central, el ruido que resulta de la suma de efectos de muchas fuentes tiende a tener una distribución gaussiana. Por esto el ruido gaussiano a menudo se usa para caracterizar la porción analógica de los canales de comunicación.

Las técnicas de modulación y demodulación se diseñan fundamentalmente para reducir los efectos del ruido gaussiano.

El segundo tipo de ruido, el impulsivo, que se encuentra en el canal, se caracteriza por largos intervalos de quietud seguidos por *bursts* (ráfagas) de pulsos de ruido de amplitud. El ruido impulsivo se debe a los transitorios de conmutación, descargas de encendido, etc. Su caracterización es mucho más difícil que la del ruido gaussiano. Además, las técnicas de modulación analógicas no son tan convenientes como los métodos de codificación digital para encarar con los fenómenos de ruido impulsivo. Por estas razones los efectos de ruido impulsivo se incluyen en el modelo de la porción discreta del canal y el ruido gaussiano se incluye en el modelo de la porción analógica del canal.

La porción analógica del canal de comunicación puede modelarse tal como se ve en la figura II.14. La entrada al canal es un proceso aleatorio $X_c(t)$, que consiste en la colección de todas las formas de onda generadas por el modulador. El ancho de banda de $X_c(t)$ y del canal se supone que es B Hertz. El ruido aditivo $n(t)$ en la salida del canal tiene promedio cero y es ruido blanco gaussiano de banda limitada. La capacidad de esta porción del canal se encuentra maximizando la tasa de transmisión de información respecto a la distribución de $X_c(t)$.

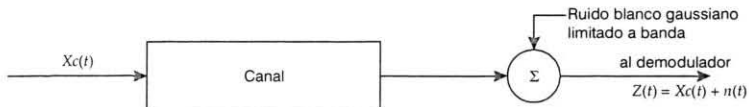


Figura II.14. Porción analógica del canal de comunicación

Puesto que la formulación de este problema es similar a la que se usa para canales discretos en términos de $H(X_c)$ y $H(X_c/Z)$, parece necesaria una optimización.

El resultado, sin embargo, tiene una forma simple; nosotros estableceremos el resultado como un teorema (véase el libro de Shannon para su demostración) y discutiremos cómo puede usarse en el diseño de sistemas de comunicaciones.

II.16. EL TEOREMA DE SHANNON-HARTLEY

La capacidad de un canal continuo con un ancho de banda B y ruido blanco gaussiano limitado en banda es

$$C = B \cdot \log_2 \left(1 + \frac{S}{N} \right) \text{ bits/seg} \quad (\text{II.34})$$

donde S y N son la potencia media de la señal y la potencia de ruido respectivamente a la salida del canal ($N = N_0 \cdot B$) si la densidad espectral de potencia de ruido de doble lado es $N_0/2$ (watts/seg).

Este teorema conocido como el teorema de Shannon/Hartley es de fundamental importancia y tiene dos implicaciones importantes para ingenieros en sistemas de comunicaciones, claramente analizadas por Shanmugan.

Primero, da el límite superior que puede alcanzarse en la forma de tasa de transmisión de datos confiables sobre un canal gaussiano, por lo tanto, un diseñador de sistemas siempre trata de optimar su sistema para tener una tasa de datos tan cercana a C como la expresada en la ecuación (II.34) con una aceptable tasa de error.

La segunda implicación del teorema de Shannon-Hartley trata con el intercambio de ancho de banda y la relación señal a ruido. Para aclarar este aspecto del teorema supongamos que queremos transmitir datos a una tasa de 10 000 bits/seg sobre un canal con un ancho de banda $B = 3\,000$ Hz. Para transmitir datos a una tasa de 10 000 bits/seg necesitamos un canal con una capacidad de por lo menos 10 000 bits/seg. Si la capacidad del canal es menor que la tasa de datos, la transmisión sin errores no es posible. Así como $C = 10\,000$ bits/seg, podemos obtener que los requerimientos de S/N del canal que son

$$\frac{S}{N} = 2^{(C/B)} - 1 = 2^{3.333} - 1 \cong 9$$

Para el mismo problema, si nosotros tenemos un canal con un ancho de banda de 10 000 Hz necesitamos una S/N de 1.

Así, una reducción de ancho de banda de 10 000 Hz a 3 000 Hz da como resultado un aumento en la potencia de la señal de 1 a 9.

Otro interesante aspecto del teorema de Shannon-Hartley es la compresión de ancho de banda. Para ilustrar este caso nos preguntamos: ¿es posible cuantificar y transmitir una señal cuyo espectro se extiende hasta una frecuencia f_m sobre un canal con un ancho de banda menor que f_m ?

La contestación es sí, y podemos justificarla como sigue: supongamos que muestreamos la señal analógica a una tasa de $3f_m$ muestras/seg (es decir a 1.5 veces la tasa de Nyquist por ejemplo) y cuantificamos el valor de la señal en uno de M posibles niveles, entonces la tasa de datos de la señal cuantificada es $3f_m \log_2 M$ (bit/seg).

Si el ancho de banda del canal es B , por una apropiada elección de la potencia de la señal podemos lograr una capacidad C mayor que $3f_m \log_2 M$. Por ejemplo, con $M = 64$ y un ancho de banda del canal igual a la mitad del ancho de banda de la señal, nosotros necesitaríamos una S/N de 109 dB para ser capaces de transmitir la señal cuantificada con una pequeña probabilidad de error. Así, una comprensión de ancho de banda de dos es posible si nosotros podemos mantener una S/N de 109 dB (valor no práctico) en la salida del canal.

Suponemos que la distorsión de la señal debida al muestreo y a la cuantificación es despreciable.

El teorema de Shannon-Hartley indica que un canal sin ruido tiene una capacidad infinita; sin embargo, cuando el ruido está presente, la capacidad del canal no se aproxima a infinito cuando el ancho de banda aumenta debido a que la potencia de ruido también aumenta. La capacidad del canal llega a un límite superior finito con el aumento del ancho de banda si la potencia de la señal es fija. Nosotros podemos calcular este límite como sigue: con $N = N_0 B$ donde $N_0/2$ es la densidad espectral de potencia de ruido, tenemos

$$C = B \cdot \log_2 \left(1 + \frac{S}{N_0 B} \right) = \left(\frac{S}{N_0} \right) \cdot \left(\frac{N_0 B}{S} \right) \cdot \log_2 \left(1 + \frac{S}{N_0 B} \right) \quad (\text{II.35})$$

$$= \left(\frac{S}{N_0} \right) \cdot \log_2 \left(1 + \frac{S}{N_0 B} \right)^{N_0 B/S}$$

Recordando que $\lim_{B \rightarrow \infty} (1+x)^{1/x} = e$ y tomando $x = S/N_0 B$, en la ecuación (II.35) tenemos que

$$\lim_{B \rightarrow \infty} C = \frac{S}{N_0} \cdot \log_2 e = 1.44 \left(\frac{S}{N_0} \right) \quad (\text{II.36})$$

Un sistema de comunicación capaz de transmitir información a una tasa de $B \log_2 (1 + S/N)$ es llamado un sistema ideal. Shannon propuso la siguiente idea para tal sistema (véase Shanmugan):

Supongamos que la fuente emite M mensajes equiprobables de T seg de longitud. El sistema de comunicación ideal observa la salida de la fuente por T segundos y el mensaje se representa (codifica) por una señal en el canal elegida de una colección de M funciones muestra de ruido blanco de duración T . En la salida del canal, la señal recibida más ruido se compara con versiones almacenadas de las señales del canal. La señal que mejor se adapta a la señal de mayor ruido se supone que ha sido transmitida y el mensaje correspondiente es decodificado. La cantidad total de retardo de tiempo utilizado en observar el mensaje, transmitirlo y decodificarlo en el receptor es favorablemente T seg.

El esquema ideal usando señales ruidosas puede llevar información a una tasa aproximada a la capacidad del canal sólo cuando $T \rightarrow \infty$. Solamente en el caso límite nosotros tenemos todas las condiciones satisfechas; en esta condición límite el sistema ideal tiene las siguientes características:

- La tasa de información $\rightarrow B \log_2 (1 + S/N)$
- La tasa de error $\rightarrow 0$
- Las señales transmitidas y recibidas tienen las características de ruido blanco y gaussiano limitado en banda
- Como $T \rightarrow \infty$ el número de señales $M \rightarrow \infty$ el retardo de la codificación también $\rightarrow \infty$.

Es obvio por la discusión precedente que un sistema ideal no puede realizarse en la práctica. Más que tratar de diseñar un sistema usando un gran número de señales analógicas, nosotros usamos un pequeño número en los sistemas prácticos, que conduce a una probabilidad de error Pe no nula. La tasa de datos y la probabilidad de error definen un canal discreto cuya capacidad C' será menor que $B \log_2 (1 + S/N)$. Por medio de este canal digital tratamos de lograr una tasa de datos aproximándose a C' con una probabilidad de error cercana a cero, usando codificación de control de error

digital. Así, en los sistemas prácticos rara vez tratamos de lograr la máxima tasa teórica de transmisión de información sobre la porción analógica del canal. Se trata de mantener esta parte del sistema razonablemente simple. En la parte digital del sistema tratamos de lograr una tasa que se aproxime a la capacidad de la porción discreta del canal, puesto que la codificación digital es más fácil de implementar.

En los siguientes capítulos veremos algunos esquemas para transmitir información digital por medio de un canal de comunicación analógica; para cada tipo de esquema se obtendrá el de ancho de banda requerido, relación S/N de salida, y la tasa de datos. Estos parámetros definirán los de la porción discreta del canal. También veremos más adelante algunos métodos de codificación para controlar el error y que nos permitirán transmitir información sobre el canal discreto a una tasa cercana a su capacidad, con una baja probabilidad de error. Nótese que la ecuación (II.34) es para un canal gaussiano. Esta limitación no disminuye la importancia de la ley de Shannon/Hartley por las siguientes razones:

Primero, muchos de los canales físicos generalmente son, o por lo menos aproximadamente, gaussianos. Segundo, se ha demostrado que el resultado obtenido para el canal gaussiano provee un límite inferior sobre el desempeño de un sistema que opera sobre un canal no gaussiano. Es decir, que si un particular codificador/decodificador produce una probabilidad de error P_e sobre el canal gaussiano, otro codificador/decodificador puede diseñarse para un canal no gaussiano que dé una menor probabilidad de error.

A continuación daremos algunos ejemplos para ilustrar el uso de la capacidad del canal en el diseño de sistemas de comunicaciones.

Ejemplo II.34

Se tiene un canal pasabanda continuo de acuerdo con el modelo de la figura. Si la potencia de la señal es 10 watts y la densidad espectral de ruido es $N_0 = 1$ mwatt/Hz, grafique la capacidad como una función del ancho de banda del canal,



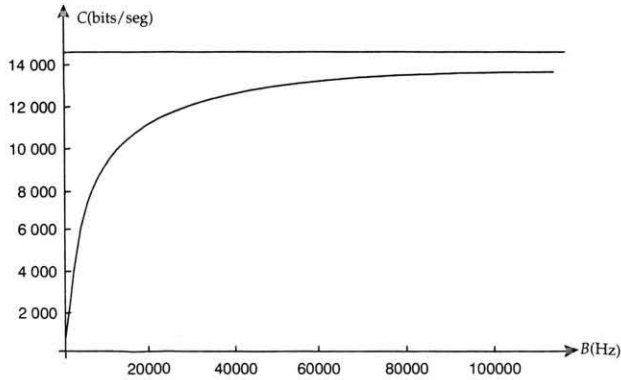
Respuesta

$$C = B \cdot \log_2 (1 + S/N) = B \cdot \log_2 \left(1 + \frac{10^4}{B} \right)$$

El siguiente cuadro de valores se obtiene

B	C
1 000 Hz	3 459 bits/seg
5 000	7 924
10 000	10 000
50 000	13 152
100 000	13 750
∞	$(S/N_0) \log_2 e = 1.44 \cdot 10\,000 = 14\,400$

Graficando estos valores



Con el objeto de comprender mejor la ley de Shannon-Hartley, escribiremos la ecuación (II.35) de una manera diferente, tomando en cuenta que la potencia de la señal es la energía por bit dividida entre la duración de un bit, $S = E_b/T$. Como la tasa de información es $R = (\log_2 M)/T$, siendo M el número de señales equiprobables, entonces

$$S = \frac{E_b}{T} = E_b \cdot \frac{R}{\log_2 M}$$

que reemplazada en la ecuación (II.36)

$$\lim_{B \rightarrow \infty} C = 1.44 \cdot \frac{S}{N_0} = 1.44 \cdot \frac{E_b \cdot R}{N_0 \cdot \log_2 M}$$

$$\text{para } R = C \Rightarrow \frac{E_b \cdot R}{N_0 \cdot \log_2 M} = \frac{1}{1.44} = -1.60 \text{ dB}$$

Así que para $\frac{E_b}{N_0 \cdot \log_2 M}$ mayor que -1.6 dB podemos comunicarnos con error despreciable.

Una comunicación confiable no es posible para más bajas relaciones S/N .

Ejemplo II.35

Calcule la capacidad de un canal pasabajas con un ancho de banda utilizable de 3 000 Hz y una $S/N = 1\,000$ en el canal de salida. Suponga que el ruido del canal es gaussiano y blanco.

Respuesta:

La capacidad C la da la ecuación

$$C = B \cdot \log_2 \left(1 + \frac{S}{N} \right) = (3\,000) \cdot \log_2 (1 + 1\,000) = 30\,000 \text{ bits/seg}$$

Los valores de los parámetros utilizados en este ejemplo son típicos de líneas telefónicas de voz estándar. La máxima tasa de datos sobre estos canales está en el orden de 10 000 bits/seg. Pueden lograrse tasas mayores pero eso requiere esquemas de modulación y demodulación muy complejos.

Ejemplo II.36

Se usa una terminal TRC para proporcionar datos alfanuméricos a una computadora. El TRC se conecta a la computadora por medio de una línea telefónica que tiene un ancho de banda útil de 3 000 Hz y una S/N de salida de 10 dB. Supongamos que la terminal tiene 128 caracteres y que los datos enviados desde la terminal consisten en secuencias independientes de caracteres equiprobables.

- a) Encuentre la capacidad del canal suponiendo el ruido blanco y gaussiano.
- b) Encuentre la tasa máxima teórica en la cual los datos pueden transmitirse.

Respuesta:

- a) La capacidad se da por:

$$C = B \cdot \log_2 \left(1 + \frac{S}{N} \right) = (3\,000) \cdot \log_2 11 = 10\,378 \text{ bits/seg}$$

- b) Contenido de información promedio/carácter:

$$H = \log_2 128 = 7 \text{ bits/carácter}$$

y la tasa de información promedio de la fuente $R = r \cdot H$. Para transmisión sin error necesitamos

$$R = r \cdot H < C \Rightarrow 7r < 10\,378 \Rightarrow r < 1\,482$$

De aquí que la tasa máxima a la que los datos pueden transmitirse sin errores es 1 482 caracteres por segundo.

II.17. PROBLEMAS

- II.1. Diga cuántos bauds y cuántos bits de información lleva la transmisión de un tono senoidal de 1 000 ciclos por segundo.
- II.2. Una forma de onda que tiene una función de densidad de amplitud gaussiana con valor medio cero y varianza σ^2 es muestreada con 100 muestras por segundo.
Las muestras x_i se cuantifican de acuerdo con el cuadro que aparece a continuación.

- a) Calcule la entropía en la salida del cuantificador.
- b) Calcule la tasa de información en bits/seg en la salida del cuantificador.

Entrada cuantificador	Salida cuantificador
$-\infty < x_i < -\sigma$	m_0
$-\sigma < x_i < 0$	m_1
$0 < x_i < \sigma$	m_2
$\sigma < x_i < \infty$	m_3

II.3. La señal de una fuente definida por la función de densidad de probabilidad

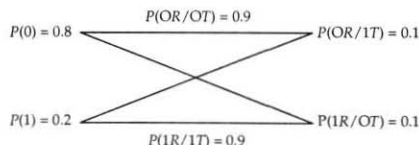
$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-(0.5x^2)}$$

es la entrada a un convertidor A/D de tres bits. El convertidor A/D es definido por el siguiente cuadro:

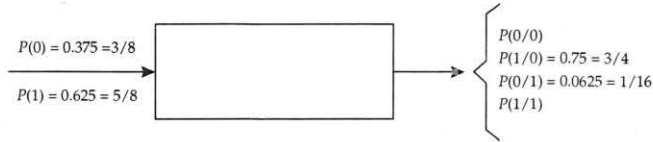
Salida	Entrada
000	$x < -2$
001	$-2 < x < -1$
010	$-1 > x < -0.5$
011	$-0.5 < x < 0$
100	$0 < x < 0.5$
101	$0.5 < x < 1$
110	$1 < x < 2$
111	$2 < x$

Determine la entropía de la fuente.

- II.4. Determine la información asociada al sacar una carta de un mazo de 52 barajas. ¿Y si saca un par de cartas?
- II.5. En una industria hay 25% de ingenieros, de ellos 60% toman café. Si 70% de los empleados toman café, qué información se recibe si el que toma café es un ingeniero.
- II.6. Supongamos que una tasa de voz normal es de 300 palabras por minuto y que la palabra promedio contiene seis caracteres considerando que son estadísticamente independientes dentro de una palabra. Calcule la tasa de información en bits/seg.
- II.7. Se dispone del siguiente canal de comunicación: recibir un 1 a la salida significa la detección de un avión enemigo; un 0 significa condiciones normales. Si se recibe un 1 se quiere calcular $P(1T/1R)$, es decir, la probabilidad de que habiendo recibido un 1 en la salida se haya transmitido un 1 en la entrada. Calcule además $P(OT/OR)$ y $P(1T/OR)$.

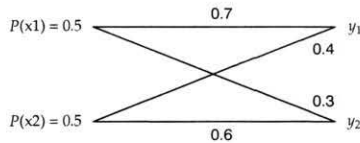


II.8. Para el esquema de la figura calcule la entropía.



II.9. Demuestre que $H(S^n) = nH(S)$; antes de resolver este ejercicio es conveniente que revise la sección II.13 (extensión de una fuente).

II.10. Para el canal binario de la figura halle $P[x, y]$



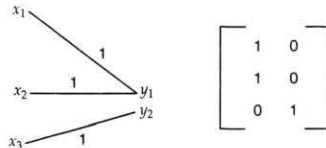
II.11. Una fuente binaria tiene salidas A y B con probabilidades de 0.75 y 0.25 respectivamente, y es la entrada a un canal definido por la matriz

$$P[(y/x)] = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$$

Determine la información mutua.

II.12. Calcule la capacidad de un canal con entradas y salidas estadísticamente independientes.

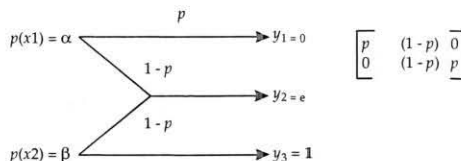
II.13. Un canal se describe por



$$\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Halle la capacidad.

II.14. Encuentre la capacidad de canal de la siguiente figura: el valor $y_2 = e$ indica que no hay salida en la entrada correspondiente.



$$\begin{bmatrix} p & (1-p) & 0 \\ 0 & (1-p) & p \end{bmatrix}$$

II.15. Encuentre la capacidad del canal de la figura.

P	Q	0	0
Q	P	0	0
0	0	P	Q
0	0	Q	P

- II.16. Una fuente tiene cuatro mensajes: m_1, m_2, m_3, m_4 , con probabilidades 0.1, 0.2, 0.3, 0.4. La salida de la fuente se codifica usando la técnica de Shannon-Fano. Determine las palabras del código y la eficiencia. Suponga que la segunda extensión de la fuente es codificada. ¿Cuál es el aumento en la eficiencia?
- II.17. Una fuente tiene cinco mensajes: m_1, m_2, m_3, m_4, m_5 , con probabilidades 0.1, 0.2, 0.2, 0.1, 0.1. La segunda extensión de la fuente se codifica con un alfabeto de tres símbolos (0, 1, 2). Determine las palabras del código.
- II.18. Una fuente tiene ocho mensajes equiprobables. Determine el código de Shannon-Fano y la eficiencia.
- II.19. Igual que el anterior, pero con nueve diferentes mensajes de salida equiprobables.
- II.20. Una fuente emite A y B con $P(A) = 0.8$ y $P(B) = 0.2$ a una tasa de $4/3$ símbolos/seg; el canal tiene una capacidad de 1 símbolo/seg. ¿La salida de la fuente puede pasar por este canal? Si es posible, encuentre una codificación adecuada.
- II.21. Una fuente binaria tiene salidas A y B con probabilidades 0.75 y 0.25. Determine un código para la tercera extensión y luego para la cuarta. Repita el problema suponiendo que $P(A) = 0.5$ y $P(B) = 0.4$.
- II.22. Una imagen de TV consta de 625 líneas con 600 elementos por línea y 32 niveles equiprobables de brillo por elemento. Si hay 25 imágenes/seg y se quiere transmitir la señal por un canal de $S/N = 50$ dB, ¿cuál es el ancho de banda necesario para el canal?
- II.23. Se quiere transmitir una señal eléctrica de 1 000 pulsos binarios equiprobables por segundo y recibirlos con una baja probabilidad de error. ¿Es posible usar un medio de transmisión con un ancho de banda $B = 1$ c/s? ¿Es posible calcular la S/R ?
- II.24. Un canal discreto sin ruido tiene $r = 10^5$. Analice la posibilidad de reemplazarlo por un canal continuo de $B = 8$ KHz y $S/N = 31$.
- II.25. Encuentre el tiempo mínimo requerido para transmitir 600 dígitos hexadecimales sobre el canal continuo del ejercicio 24.
- II.26. Estime $M(T)$ para un canal telefónico típico con $B = 3$ KHz, $S/N = 30$ dB y $T = 3$ minutos. Recuerde que $M(T)$ es la cantidad de mensajes posibles.
- II.27. Calcule S/N en dB para un canal telefónico si se quieren transmitir dígitos binarios a una tasa $R = 2$ B bits/seg, y para $R = 4$ B.

BIBLIOGRAFÍA

- [1] Shannon, C., *Mathematical Theory of Communication*, Chicago, University of Illinois Press.
- [2] Abramson, *Teoría de la información y codificación*, México, Paraninfo.
- [3] Carlson, B., *Sistemas de comunicaciones*, México, McGraw-Hill.
- [4] Wozencraft and Jacobs, *Principles of Communications Engineering*, John Wiley.
- [5] Shanmugan, *Digital and Analog Communication Systems*, John Wiley.
- [6] Ziemer y Tranter, *Principles of Communications*, Londres, Houghton Mifflin.

CAPÍTULO III

MUESTREO

III.1. ALGUNAS DEFINICIONES

A

NTES DE ABUNDAR en detalles sobre los sistemas de comunicaciones digitales, es conveniente definir algunas cantidades que serán de utilidad más adelante

III.1.1. Señales y fuentes analógicas

Aunque algunas de estas definiciones ya fueron esbozadas en la introducción del capítulo I, las repetiremos aquí para mayor claridad. Una *fente analógica* de información produce una salida que puede tomar cualquier valor dentro de un intervalo continuo de valores a cualquier tiempo. De manera similar, una *señal analógica* es una forma de onda eléctrica que puede tomar cualquier amplitud posible dentro de un intervalo continuo de amplitudes a cualquier tiempo. La presión de sonido de una orquesta tocando música es un ejemplo de una fuente analógica, mientras que el voltaje de un micrófono respondiendo a las ondas de sonido representa una señal analógica.

III.1.2. Señales y fuentes digitales

Una *fente digital* es aquella que tiene una salida que puede tomar únicamente un valor de un conjunto finito de valores a cualquier tiempo. La mayoría de las fuentes en la naturaleza son analógicas, sin embargo, cuando se combinan con algún dispositivo manufacturado puede dar como resultado una fuente digital. Por ejemplo, la temperatura es una cantidad analógica pero cuando se combina con un termostato cuya salida solamente es encendido y apagado, la combinación puede considerarse como una fuente digital.

Una *señal digital* se define como una forma eléctrica que únicamente toma una de un conjunto finito de amplitudes en cualquier tiempo. Si el termostato se diseña para que dé un cierto valor de voltaje cuando está encendido y cero cuando esté apagado, esta salida se le puede considerar como una señal digital, también conocida como mensaje. En general, la palabra mensaje frecuentemente se utiliza para referirse a las salidas de la fuente, por lo que una fuente digital puede decirse que entrega uno de un conjunto finito de mensajes en cualquier tiempo. En la figura III.1 se muestran la forma de onda de una señal analógica y de una señal digital; aquí vale la pena aclarar que una señal digital no necesariamente puede tomar sólo dos amplitudes distintas (como en este ejemplo), sino en general un número finito de amplitudes. Entendemos por un conjunto finito aquel que podemos contar.

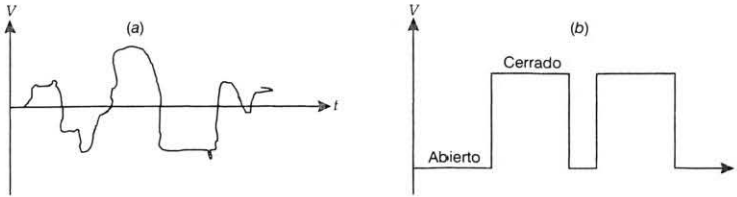


Figura III.1. a) Forma de onda de una cierta señal analógica, b) forma de onda de una señal digital

III.1.3. Clasificación de señales

Suponemos en todo este desarrollo que las señales de información analógicas y digitales definidas previamente son de banda base, a menos que se indique otra cosa. Una señal de *banda base* es aquella que tiene a la mayoría de sus componentes espectrales agrupados en una banda de frecuencia alrededor (o cercana) de la frecuencia cero. El término *pasa bajas* también se utiliza algunas veces para referirse a señales de banda base. Todas las señales *pasa bajas* prácticas tienen una frecuencia arriba de la cual sus componentes espectrales pueden considerarse despreciables. Denominaremos esta frecuencia como el *espectro extendido* de la señal y la identificaremos como W_f (sus unidades son radianes/seg).

Una señal *pasa banda* es aquella que tiene a la mayoría de sus componentes espectrales agrupados en una banda de frecuencia desplazada por un incremento significativo de la frecuencia cero. La mayoría de las señales de información son de banda base, raramente son *pasa bandas*, pues por lo regular son el resultado de un proceso de modulación.

En la figura III.2 se ilustran los conceptos anteriormente definidos; la figura III.2a corresponde al espectro de una señal de banda base (*pasa bajas*) con un espectro extendido W_f , la figura III.2b ilustra el espectro de una señal *pasa banda* y su correspondiente espectro extendido.

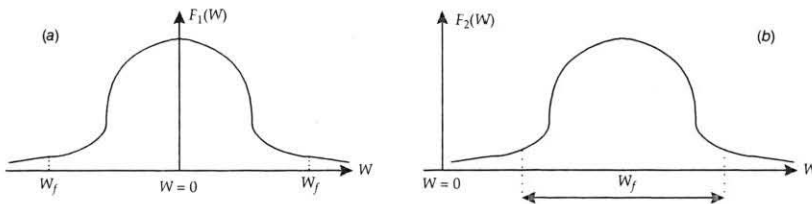


Figura III.2. a) Espectro extendido de una señal de banda base (*pasa bajas*), b) espectro extendido para una señal *pasa banda*

III.1.4. Modulación

La mayoría de las fuentes de información se caracterizan por señales de información de banda base; sin embargo, las formas de onda en banda base no se transmiten eficazmente por métodos de radio, de un punto a otro. Por otro lado, las formas de onda *pasa bandas* fácilmente pueden transmitirse por radio; por lo tanto, un propósito básico de la modulación es desplazar el mensaje a una banda

de frecuencias mayor para mejor radiación. Algunos sistemas de modulación involucran cambios en la amplitud, frecuencia o fase (o combinación de ellas) de una señal portadora en función del mensaje. Cuando el mensaje es digital nos referimos a esas operaciones como *manipuleo* (*keying*); por ejemplo, manipuleo por corrimiento de fase (PSK).

III.2. DIAGRAMA A BLOQUES DE UN SISTEMA DE COMUNICACIÓN DIGITAL

Un *sistema de comunicación digital* (SCD) es aquel que transmite información usando únicamente un conjunto finito de mensajes o símbolos. Un sistema de comunicación típico está formado por una estación de transmisión, una estación de recepción y un medio de conexión al que se le denomina *canal*. Estas funciones básicas son adecuadas para una operación en una sola dirección; una comunicación bidireccional requiere que cada estación tenga tanto un transmisor como un receptor. Puesto que cada dirección de comunicación es similar, únicamente describiremos la operación en un sentido.

III.2.1. Subsistema de transmisión

En la figura III.3 se aprecia un diagrama a bloques de las principales funciones que pueden estar presentes en un sistema de comunicación digital; puesto que todo el sistema es digital, el subsistema de transmisión puede aceptar directamente señales digitales; además, también puede trabajar con señales analógicas, pero éstas deben ser convertidas a un formato digital por medio de un convertidor analógico a digital (CAD). El proceso de conversión A/D abarca el muestreo periódico de la forma de onda analógica (la teoría que analizaremos más adelante en este capítulo) y la cuantización de las muestras.

El proceso de cuantización es un proceso de redondeo de la amplitud de las muestras, donde cada muestra toma el valor más cercano, seleccionado de un conjunto discreto de amplitudes. El resultado es una señal digital que es compatible con el sistema digital. El proceso de redondeo sufrido por las muestras en la etapa de cuantización, limita la exactitud con que la señal analógica puede reconstruirse en el receptor; los métodos de reconstrucción los analizaremos posteriormente en este mismo capítulo, aunque los principios de cuantización los explicamos en el capítulo IV.

Como ya dijimos, la salida actual del convertidor A/D en el punto A del diagrama es un nivel discreto de voltaje. Para describir la siguiente función, el *codificador de la fuente*, es útil simplificar la imagen de la señal digital sin importar su origen, la cual tiene uno de un conjunto discreto de símbolos (niveles) en cualquier tiempo.

El propósito general del codificador de la fuente es convertir efectivamente cada símbolo discreto en una representación digital adecuada, algunas veces binaria. Por ejemplo, supongamos que un mensaje digital puede tener cinco diferentes símbolos (niveles), denotados por m_1, m_2, m_3, m_4 y m_5 , tal como se muestra en la figura III.4b. Estos niveles pueden representarse por una secuencia de dígitos binarios 0 y 1, mostrados en la segunda columna de la figura III.4a. Cada secuencia tiene tres dígitos; cada dígito en la secuencia puede ser 0 o 1. Claramente hay $2^3 = 8$ palabras de código, por lo cual una representación de tres dígitos binarios puede soportar hasta ocho símbolos.

Para propósitos de comparación usaremos ahora una representación ternaria con una longitud de palabra de dos dígitos (este código se muestra en la tercera columna del cuadro); en este caso los dígitos que se pueden utilizar son 0, 1 y 2, lo que nos da $3^2 = 9$ posibles palabras de código ternarias, por lo que se pueden representar hasta nueve símbolos por medio de este código ternario.

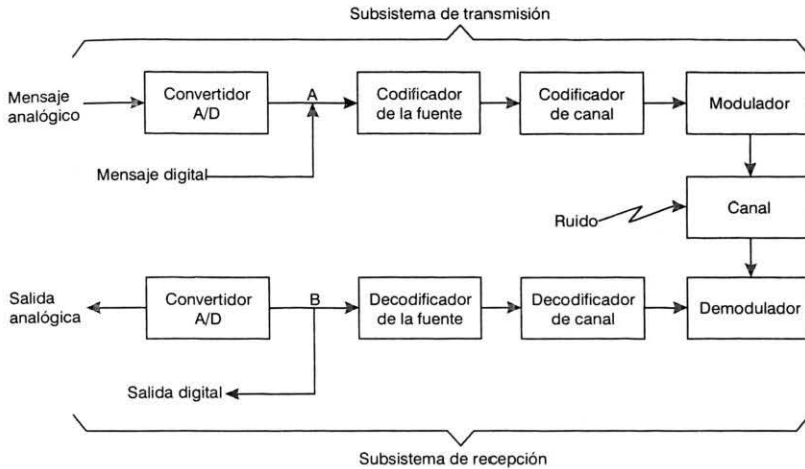
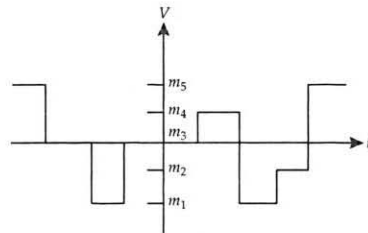


Figura III.3. Diagrama funcional a bloques de un sistema de comunicación digital

Se dice que los mensajes digitales poseen *redundancia* cuando sus símbolos no son igualmente probables o no son estadísticamente independientes. Muchas fuentes de mensajes prácticos, por ejemplo, el lenguaje español, tienen redundancia: *tra_e d_ comp_n_e_ _s_a o_ci_*. Uno de los objetivos del codificador de la fuente es remover la redundancia. Entre más eficiente es el codificador más redundancia se remueve, lo que permite que se *reduzca* el número de dígitos binarios para representar el mensaje. En el capítulo II explicamos algunos métodos de codificación de fuente.

Nivel (símbolos)	Binario natural	Ternario
m_5	000	00
m_4	001	01
m_3	010	02
m_2	011	10
m_1	100	11

(a)



(b)

Figura III.4. a) Codificación en código binario y ternario de la señal digital, b) señal digital con cinco símbolos

En algunos sistemas donde no está presente la función de codificación de canal, la salida del codificador de la fuente se convierte directamente a una forma apropiada (con la función del modulador) para transmitir sobre el canal. El ruido y la interferencia sumados a la forma de onda, causan que en el demodulador del subsistema de recepción se produzcan errores en el proceso de recuperar (determinar) el mensaje enviado por el transmisor. Al incluir la función del *codificador de canal*, se pueden reducir los efectos de los errores causados por el canal; el codificador de canal hace esto posible sumando redundancia controlada a la señal digital, de una manera conocida tal que los errores pueden reducirse. La codificación de canal se tratará en el capítulo IV.

La función del modulador es convertir la señal digital a una forma apropiada para transmitirla sobre el canal; este bloque se utiliza principalmente cuando la señal digital se transmite sobre canales analógicos (por ejemplo la red telefónica), y los esquemas de modulación que se utilizan principalmente son *PSK*, *FSK*, *ASK* y modulación multisímbolo, que analizaremos en el capítulo V.

III.2.2. Canal de comunicación

A lo largo de este libro, supondremos que el canal es lineal e invariable en el tiempo, a menos que se indique otra cosa, y se considerará que el canal tendrá un ancho de banda infinito y que el ruido sumado en el canal es blanco y gaussiano con un espectro de densidad de potencia constante igual a $N_0/2$ aplicable de $-\infty < w < \infty$.

III.2.3. Subsistema de recepción

Las funciones hechas en este subsistema son las operaciones inversas a las efectuadas en el subsistema de transmisión. El demodulador reproduce lo mejor posible una versión de la salida que fue producida por el codificador de canal del transmisor. La salida del demodulador puede tener errores ocasionales causados por el ruido en el canal. Parte de la optimización de diversos sistemas de comunicación digital se centra en minimizar los errores producidos en el demodulador. Los sistemas de banda base óptimos los trataremos en el capítulo IV.

El propósito del decodificador de canal es reconstruir lo mejor posible la salida generada por el codificador de la fuente en el transmisor. Es aquí donde puede usarse la redundancia controlada, insertada por el codificador de canal, para detectar y corregir algunos errores en la salida del demodulador causados por el canal. La decodificación de canal es tema del capítulo V.

Cabe aclarar que las funciones que pueden estar presentes en un sistema de comunicación digital no siempre reflejan la realización real del sistema. Algunos sistemas prácticos no siempre siguen la secuencia ni incluyen todos los bloques mostrados en la figura III.3. Por ejemplo, en el capítulo IV se analizarán varios sistemas que pueden aceptar una señal analógica directamente y producir la forma de onda que se transmitirá al canal en una sola operación. Estos sistemas no usan ni codificación ni decodificación de canal.

III.3. MUESTREO IDEAL

Cuando un mensaje analógico se transmite sobre un sistema de comunicación, la totalidad del mensaje es enviado durante todo el tiempo. Para enviar la misma señal analógica sobre un sistema de comunicación digital, se requiere que únicamente sus *muestras* se transmitan a intervalos periódicos. Puesto que el receptor sólo puede recibir muestras del mensaje, éste debe tratar de reconstruir el mensaje original a partir de las muestras para todo el tiempo. Hay métodos con los que se logra este objetivo y que incluyen la teoría de muestreo, que analizaremos en este capítulo.

A primera vista puede asombrarnos que sólo muestras del mensaje y no la señal completa sean necesarias para describir toda la información de la señal. Sin embargo, a pesar de lo sorprendente como pueda ser, se puede encontrar que, en algunas condiciones razonables, un mensaje puede reconstruirse *exactamente* a partir de sus muestras. Para lograr este propósito se demostrarán algunos principios de muestreo (teoremas) que pueden aplicarse tanto a señales determinísticas como a

señales aleatorias (ruido); estos principios establecen los fundamentos que hacen posibles muchos de los sistemas digitales que estudiaremos en los siguientes capítulos.

Una de las aplicaciones más importantes que pueden realizarse a partir del muestreo es que es posible intercalar en el tiempo muestras de diferentes señales de información. Esto nos permite realizar un proceso de *multiplexaje por división en el tiempo*, análogo al multiplexaje en frecuencia. Con la disponibilidad de circuitos de conmutación de alta velocidad, actualmente se prefiere cada vez más el multiplexaje por división en el tiempo que el multiplexaje de frecuencia.

III.3.1. Teorema de muestreo para señales pasa bajas

Las preguntas básicas que surgen cuando muestreamos una señal son:

- ¿Cuál debe ser la frecuencia de muestreo?
- ¿Se pierde información por el muestreo?
- ¿Es posible recuperar la señal original sin distorsión a partir de las muestras, y cómo?

Estas preguntas se contestan por el teorema de muestreo, que establece lo siguiente:

Sea $f(t)$ una señal pasa bajas, limitada en banda, es decir que no tiene componentes de frecuencia mayores que W_f rad/seg. Si los valores de $f(t)$ se determinan a intervalos regulares separados por tiempos $T_s \leq \pi/W_f$ segundos, esto es, la señal es periódicamente muestreada cada T_s segundos. Entonces estas muestras $f(nT_s)$ determinan únicamente a la señal, y ésta puede ser reconstruida a partir de estas muestras sin distorsión.

A este teorema algunas veces se le llama teorema de muestreo uniforme, debido a que las muestras están uniformemente espaciadas; la condición de muestreo también puede expresarse en términos de la frecuencia en vez del periodo, como se establece en el teorema, simplemente tomando el inverso de la condición, con lo que se obtiene:

$$T_s \leq \frac{\pi}{W_f} \Rightarrow \frac{1}{T_s} \geq \frac{W_f}{\pi}, \quad W_f = 2 \cdot \pi \cdot f_{\max}$$

$$f_s \geq \frac{2 \cdot \pi \cdot f_{\max}}{\pi} \Rightarrow f_s \geq 2 f_{\max} \quad (\text{III.1})$$

lo que determina que la frecuencia de muestreo debe ser mayor o igual a dos veces la máxima frecuencia de la señal. La frecuencia de muestreo mínima $2 W_f (2 f_{\max})$ se conoce como la frecuencia de Nyquist.

La prueba de este teorema comienza suponiendo que $f(t)$ es una señal arbitraria, excepto que su transformada de Fourier $F(w)$ existe y representa una señal limitada en banda; es decir, tiene valores distintos de cero sólo en el intervalo $-W_f \leq w \leq W_f$ tal como se ilustra en la figura III.5a. Para facilidad de cálculo se supondrá que la señal de muestreo $s(t)$ es un tren de deltas de Dirac con periodo T_s ; esta señal así como su espectro se ilustran en la figura III.5b; en la figura III.5c se indica la señal muestreada, obtenida multiplicando la señal original $f(t)$ por la señal de muestreo; asimismo se observa el espectro de la señal muestreada, que se obtiene convolucionando en el dominio de la frecuencia los respectivos espectros de $f(t)$ y de $s(t)$.

Al analizar el espectro de la señal muestreada se observa que está formado por un tren de espectros de la señal original centrados en múltiplos de la frecuencia de muestreo; a partir de esto se deduce que para recuperar la señal original (recuperar el espectro de la figura III.5a), sólo es necesario pasar las muestras por un filtro pasa bajas ideal, tal como se observa en la figura III.5c.

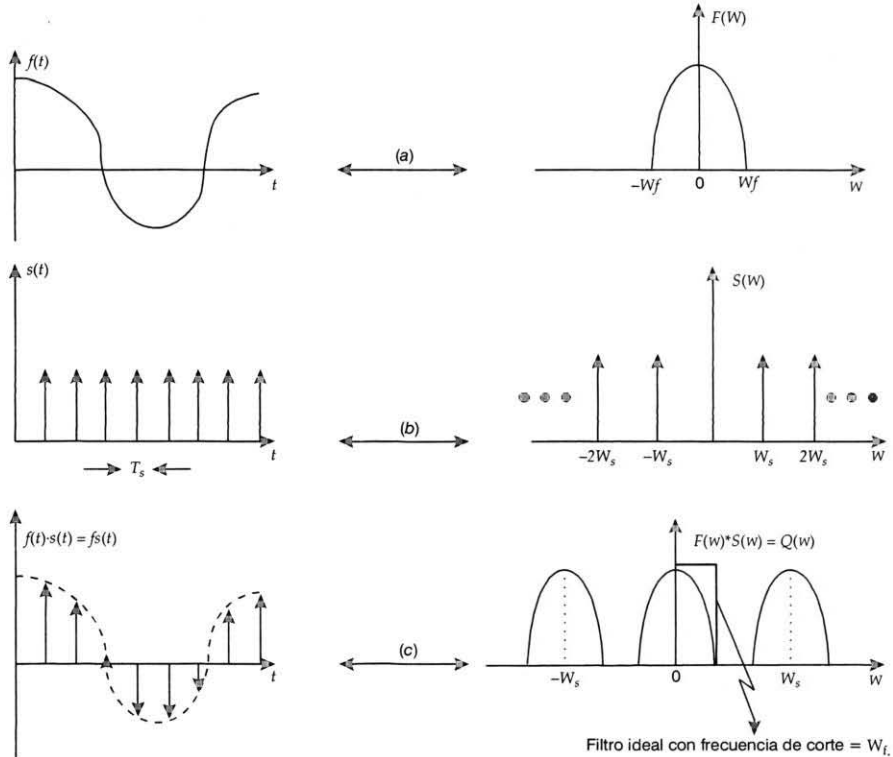


Figura III.5. Análisis gráfico del muestreo ideal

Primero vamos a demostrar que las muestras de $f(t)$ tomadas a intervalos de tiempo regulares y a una frecuencia $F_s \geq 2f_{\text{máx}}$, determinan totalmente a $f(t)$. Es decir, probaremos que no se pierde información por el muestreo. El espectro de la señal muestreada puede desarrollarse en términos de una serie de Fourier suponiendo que $Q(w)$ es periódico tal como se muestra en la figura III.5. Si el periodo de repetición es $w_s \geq 2W_f$, es obvio que no habrá traslapamiento entre los componentes espectrales; sin traslapamiento la serie de Fourier de $Q(w)$ será igual a $F(w)$ para $|w| \leq W_f$, o sea, se cumple que:

$$Q(w) = F(w) \quad \text{si} \quad |w| \leq W_f \quad (\text{III.2})$$

Ahora, escribiendo $Q(w)$ en forma de serie de Fourier exponencial, tenemos que

$$Q(w) = \sum_{n=-\infty}^{\infty} C_n e^{jn 2\pi w/w_s} \quad (\text{III.3})$$

donde los coeficientes de la serie están dados por

$$C_n = \frac{1}{w_s} \int_{-\frac{w_s}{2}}^{\frac{w_s}{2}} Q(w) e^{-jn 2\pi \frac{w}{w_s}} dw \quad (\text{III.4})$$

para $n = 0, \pm 1, \pm 2 \dots$ Como para $|w| \leq W_f \Rightarrow Q(w) = F(w)$ los coeficientes los podemos escribir como:

$$C_n = \frac{2\pi}{w_s} \frac{1}{2\pi} \int_{-W_f}^{W_f} F(w) e^{-jn 2\pi \frac{w}{w_s}} dw = \frac{2\pi}{w_s} f\left(\frac{-n 2\pi}{w_s}\right) \quad (\text{III.5})$$

de la expresión anterior se deduce que los coeficientes C_n son proporcionales a las muestras de $f(t)$ en los tiempos $\frac{-2n\pi}{w_s} = -n T_s$ donde $T_s = \frac{2\pi}{w_s}$ es el periodo de muestreo.

Entonces, a partir de la ecuación (III.3) a $Q(w)$ lo podemos escribir como

$$Q(w) = \sum_{n=-\infty}^{\infty} \frac{2\pi}{w_s} f\left(\frac{-n 2\pi}{w_s}\right) e^{jn 2\pi \frac{w}{w_s}} \quad (\text{III.6})$$

lo cual es válido para toda w . Ahora, sustituyendo (III.6) en (III.1) y efectuando la transformada inversa, tenemos

$$\begin{aligned} f(t) &= \sum_{n=-\infty}^{\infty} \frac{1}{w_s} f\left(\frac{-n 2\pi}{w_s}\right) \int_{-W_f}^{W_f} e^{jn 2\pi \frac{w}{w_s}} e^{j\omega t} dw \\ f(t) &= \sum_{n=-\infty}^{\infty} \frac{1}{w_s} f\left(\frac{-n 2\pi}{w_s}\right) \int_{-W_f}^{W_f} e^{j\omega \left(t + \frac{2\pi n}{w_s}\right)} dw \\ f(t) &= \sum_{n=-\infty}^{\infty} \frac{2 W_f}{w_s} f\left(\frac{-n 2\pi}{w_s}\right) \frac{\text{sen} \left[W_f \left(t + \frac{2\pi n}{w_s} \right) \right]}{\left[W_f \left(t + \frac{2\pi n}{w_s} \right) \right]} \end{aligned}$$

haciendo $k = -n$

$$f(t) = \sum_{k=-\infty}^{\infty} \frac{2 w_f}{w_s} f\left(\frac{2 \pi k}{w_s}\right) \frac{\text{sen} \left[W_f \left(t - \frac{2 \pi k}{w_s} \right) \right]}{\left[W_f \left(t - \frac{2 \pi k}{w_s} \right) \right]} \quad (\text{III.7})$$

Esta ecuación demuestra que si conocemos las muestras de $f(t)$ podemos determinar $f(t)$ para todo tiempo.

En términos de la función *sampling*, definida como $\text{sinc}(x) = \frac{\text{sen}(x)}{x}$, y sustituyendo $T_s = \frac{2 \pi}{w_s}$, la ecuación anterior la podemos escribir como

$$f(t) = \frac{w_f T_s}{\pi} \sum_{k=-\infty}^{\infty} f(kT_s) \text{sinc} \left[W_f(t - kT_s) \right] \quad (\text{III.8})$$

Remarcamos que estos resultados han sido derivados bajo la restricción de que $w_s \geq 2 W_f$, requiriendo un intervalo de muestreo $T_s \leq \pi/W_f$, como se estableció en el teorema. Un caso especial de la ecuación anterior ocurre cuando se muestrea la frecuencia de Nyquist, es decir cuando

$$T_s = \frac{\pi}{W_f}$$

$$f(t) = \sum_{k=-\infty}^{\infty} f(kT_s) \text{sinc} \left[W_f(t - kT_s) \right] \quad (\text{III.9})$$

Esta ecuación indica que $f(t)$ está determinada por la suma de todas las muestras pesadas por el factor $(\text{sen } x)/x$, centrado alrededor de los instantes de muestreo.

Veamos otras dos variantes del teorema de muestreo que se presentan a continuación en forma de ecuaciones, aunque no las demostraremos:

$$f(t - t_0) = \frac{W_f T_s}{\pi} \sum_{k=-\infty}^{\infty} f(kT_s - t_0) \text{sinc} \left[W_f(t - kT_s) \right] \quad (\text{III.10})$$

$$f(t) = \frac{W_f T_s}{\pi} \sum_{k=-\infty}^{\infty} f(kT_s - t_0) \text{sinc} \left[W_f(t + t_0 - kT_s) \right] \quad (\text{III.11})$$

El lector puede demostrar la ecuación (III.10) como un ejercicio; el procedimiento consiste en rehacer la demostración dada anteriormente reemplazando $F(w)$ por el espectro $F(w) \cdot \exp(-jw t_0)$, lo que corresponde a la versión desplazada de la señal $f(t - t_0)$. Como un segundo ejercicio la (III.11) se obtiene directamente de la (III.10) por inspección. (¿Cómo?)

Ahora vamos a probar que podemos recuperar sin distorsión la señal original si pasamos las muestras a través de un filtro ideal con frecuencia de corte igual a W_f , como se ilustra en la figura III.6.

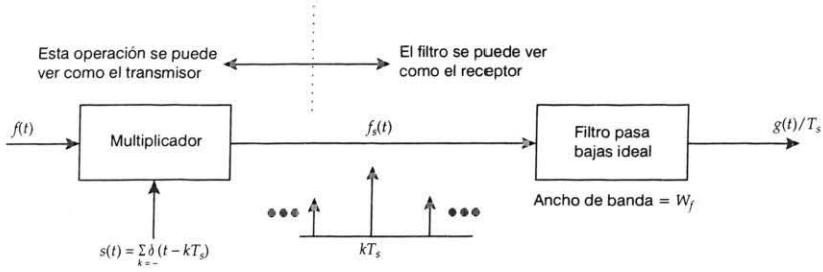


Figura III.6. Diagrama a bloques de la interpretación del teorema de muestreo

Supongamos que $f(t)$ está limitado en frecuencia a W_f . Primero vamos a calcular la transformada de Fourier de una sola muestra, ocurriendo en el k -ésimo instante de muestreo. La k -ésima muestra está dada por

$$f_n(t) = f(kT_s) \delta(t - kT_s) \quad (\text{III.12})$$

y su transformada de Fourier está dada por

$$F_n(\omega) = \int_{-\infty}^{\infty} f_n(t) e^{-j\omega t} dt = \int_{-\infty}^{\infty} f(kT_s) \delta(t - kT_s) e^{-j\omega t} dt$$

por propiedades de la delta de Dirac, el valor de la integral es

$$F_n(\omega) = f(kT_s) e^{-j\omega kT_s} \quad (\text{III.13})$$

Pero la expresión anterior no es más que la transformada de Fourier de la función impulso de amplitud $f(kT_s)$ localizada en el tiempo $t = kT_s$.

Supongamos ahora que este impulso pasa a través de un filtro pasa bajas ideal de ancho de banda W_f de amplitud unitaria y cero corrimiento de fase, cuya función de transferencia está dada por

$$H(\omega) = \begin{cases} 1, & |\omega| < W_f \\ 2, & |\omega| > W_f \end{cases} \quad (\text{III.14})$$

La respuesta al impulso $h(t)$ de este filtro ideal está dada por

$$h(t) = \frac{W_f}{\pi} \cdot \frac{\text{sen}(W_f \cdot t)}{W_f \cdot t} \quad (\text{III.15})$$

La salida $g_n(t)$ de este filtro pasa bajas a la muestra individual $f(kT_s)$ es

$$g_n(t) = \int_{-\infty}^{\infty} f_n(\tau) h(t - \tau) d\tau = \int_{-\infty}^{\infty} f(kT_s) \delta(\tau - kT_s) \frac{W_f \text{sen}[W_f(t - \tau)]}{\pi W_f(t - \tau)} d\tau$$

$$g_n(t) = \frac{W_f}{\pi} f(kT_s) \frac{\text{sen}[W_f(t - kT_s)]}{W_f(t - kT_s)} \quad (\text{III.16})$$

Entonces, aplicando el principio de superposición, la salida total del filtro es la suma de todas las contribuciones individuales de las muestras, es decir

$$g(t) = \sum_{n=-\infty}^{\infty} g_n(t) = \frac{W_f}{\pi} \underbrace{\sum_{k=-\infty}^{\infty} f(kT_s) \text{sen}[W_f(t - kT_s)]}_{f(t)} = \frac{f(t)}{T_s} \quad (\text{III.17})$$

$$\Rightarrow g(t) = \frac{f(t)}{T_s} = \text{cte} \cdot f(t)$$

la salida del filtro es proporcional a la señal original $f(t)$ en todos los instantes de tiempo. Entonces la (III.17) demuestra que la señal de entrada original $f(t)$ puede reconstruirse sin distorsión a partir de sus muestras, si éstas se pasan a través de un filtro pasa bajas ideal de ancho de banda igual a W_f .

A continuación resumiremos lo que hemos desarrollado hasta este punto. Primero, el multiplicador genera un tren de muestras instantáneas de la señal $f(t)$, con una frecuencia de muestreo $w_s = 2\pi/T_s \geq 2W_f$. Estas muestras pueden verse como la salida de un transmisor. Segundo, utilizando un filtro pasa bajas ideal con un ancho de banda igual al espectro extendido W_f de la señal $f(t)$, se obtiene una salida que está dada por la parte intermedia de (III.17); este filtro puede verse como el receptor que puede recuperar la señal original. Finalmente, aplicando el teorema de muestreo, esta salida es igual a $f(t)/T_s$, por lo que se ha demostrado que, $f(t)$ puede reconstruirse completamente sin distorsión a partir de sus muestras, utilizando un filtro pasa bajas ideal; la señal $f(t)$ se multiplica por un factor constante que únicamente altera su amplitud mas no su forma. La reconstrucción es válida para cualquier frecuencia de muestreo $w_s \geq W_f$.

El muestreo ideal de la figura III.6 no puede ser realizado; sin embargo, puede aproximarse en forma práctica utilizando como señal de muestreo un tren de pulsos muy angostos y de gran amplitud. Por fortuna, estas medidas no son necesarias, puesto que existen técnicas prácticas fácilmente realizables, que analizaremos más adelante en este capítulo.

Con el propósito de aclarar la manera como en el filtro pasa bajas se reconstruye la señal original, en la figura III.7 se muestra la gráfica de (III.17) para algunos valores de k

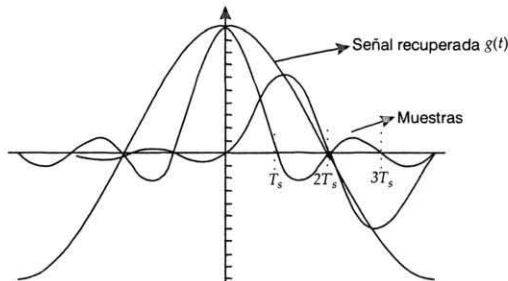


Figura III.7. Señal reconstruida a partir de sus muestras

$g_n(t)$ es máxima cuando $t=kT_s$, es decir en los instantes de muestreo donde ocurre la muestra y cero en todos los otros instantes de muestreo. Es fácil notar, a partir de la figura, que el valor de la señal recuperada en un instante diferente de los de muestreo se obtiene sumando todas las contribuciones de las muestras, y que las muestras más cercanas contribuyen en un mayor porcentaje a la definición de la amplitud en estos instantes.

El filtro ideal utilizado para reconstruir la señal es un artificio matemático; en la realidad la salida real del filtro sólo es una aproximación de la señal original. Debido a que en la práctica no se puede tener un filtro ideal, tampoco es posible muestrear a la frecuencia de Nyquist, pues provocaría que en el espectro de la señal muestreada sus componentes aparecieran uno a continuación del otro, como se observa en la figura III.8.

Para propósitos prácticos de recuperación de la señal original es necesario muestrear a una frecuencia mayor que $2W_f$. Lo que determina qué tan mayor a $2W_f$ debe ser la frecuencia de muestreo son las características del filtro pasa bajas, es decir su atenuación u orden; como un ejemplo práctico, en la figura III.9 se muestran los valores típicos utilizados para muestreo de voz.

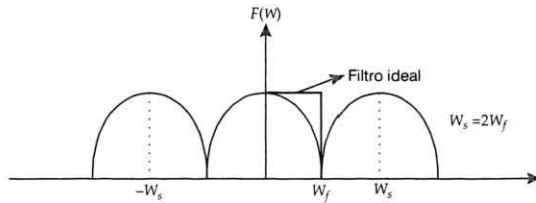


Figura III.8. Espectro de una señal muestreada a la frecuencia de Nyquist

Para aplicaciones de telefonía la voz se limita a una frecuencia máxima de 3.4 KHz, y se utiliza una frecuencia de muestreo de 8 KHz, lo que permite tener un intervalo de guarda del orden de 1 200 Hz, los filtros pasa bajas utilizados en la recuperación son típicamente de quinto o sexto orden, lo que permite tener una atenuación a la frecuencia de inicio del siguiente espectro de aproximadamente -32 a -36 dB. Esto implica que existe un compromiso entre la calidad del filtro y la frecuencia de muestreo: a mayor frecuencia de muestreo se reduce el orden del filtro y su complejidad, sin embargo se aumenta el número de muestras por segundo que se van a transmitir o procesar, lo que conlleva a necesitar un mayor ancho de banda en el canal de transmisión o mayor velocidad de cálculo. En el caso contrario se reduce tanto el ancho de banda del canal como la velocidad de cálculo pero se aumenta la complejidad del filtro de recuperación.

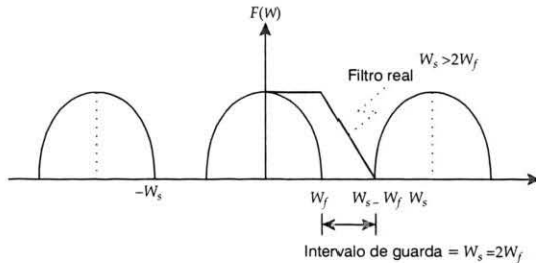


Figura III.9. Espectro de una señal muestreada a una frecuencia mayor a la de Nyquist

III.4. TRASLAPAMIENTO (ALIASING)

Examinaremos ahora el espectro de una señal muestreada idealmente $f_s(t)$. El tren de impulsos puede reemplazarse por su representación en serie de Fourier para obtener

$$f_s(t) = f(t) \sum_{k=-\infty}^{\infty} \delta(t - kT_s) = \frac{1}{T_s} \sum_{k=-\infty}^{\infty} f(t) e^{jk\omega_s t} \quad (\text{III.18})$$

la transformada de Fourier de $f_s(t)$, denotada por $F_s(w)$, es

$$F_s(w) = \frac{1}{T_s} \sum_{k=-\infty}^{\infty} F(w - k\omega_s) \quad (\text{III.19})$$

obtenida aplicando la propiedad de desplazamiento de las transformadas. $F_s(w)$ está formada por réplicas escaladas (por $1/T_s$) del espectro de la señal original centradas en todos los múltiplos de ω_s . Si $F(w)$ está limitada en banda a W_f y si $\omega_s \geq 2W_f$, entonces las réplicas no se traslaparán, lo que se requiere para que el filtro pasa bajas de la figura III.6 pueda recuperar el espectro sin distorsión de $F(w)$, que corresponde a la componente de $F_s(w)$ para $k = 0$.

Sin embargo, si $f(t)$ no está limitado en banda o si la frecuencia de muestreo ω_s no es la apropiada, puede provocar que los componentes espectrales se traslapen como se ilustra en la figura III.10. Este traslapamiento se conoce como aliasing; en la figura III.10 el aliasing se debe a que el mensaje no está limitado en banda, esta forma de aliasing puede minimizarse muestreando a una alta frecuencia de muestreo a fin de provocar que las réplicas aparezcan mucho más separadas; otra solución puede ser prefiltrar el mensaje para forzarlo a estar limitado en banda. En la figura III.11 el aliasing se produce únicamente porque la frecuencia de muestreo es menor a la frecuencia de Nyquist, la solución es aumentar ω_s a un valor apropiado.

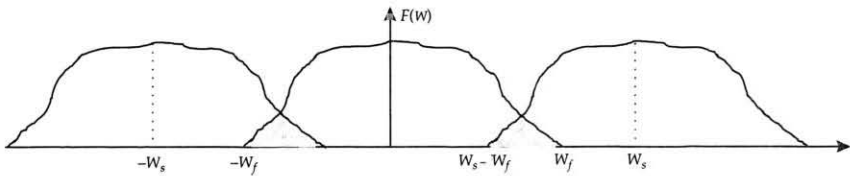


Figura III.10. Aliasing producido por una señal no limitada en banda

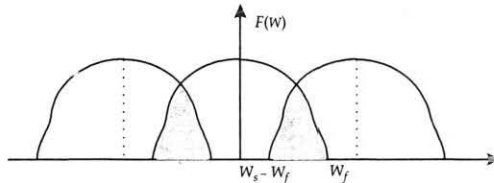


Figura III.11. Aliasing producido por muestrear a una frecuencia menor a la de Nyquist

Hasta aquí hemos supuesto que las señales a muestrear están estrictamente limitadas en banda, hecho que no se cumple para señales reales. Así por ejemplo, un espectro como el de la figura III.12a puede considerarse virtualmente limitado en banda si el contenido de componentes por encima de W_f es pequeño y prácticamente carece de importancia en relación con la información contenida desde la frecuencia cero hasta W_f . Sin embargo si muestreamos dicho espectro, suponiendo que no hay componentes para $w > W_f$, se producirá el solapamiento de espectros con la ya mencionada distorsión alineal (figura III.12b). Esto implicará, una vez realizada la etapa de reconstrucción, la aparición de componentes de frecuencia que originalmente se encontraban fuera de la banda de frecuencia considerada nominal en forma de componentes de más baja frecuencia. Así $f_a > f_{\max}$ deviene en $f_s - f_a < f_{\max}$ (figuras III.12a y b).

Este tipo de distorsión aparece siempre que alguna componente del mensaje es submuestreada y recibe el nombre de solapamiento espectral (en inglés se le denomina *foldover distortion* o *aliasing*). Por ejemplo, el proceso de filmación es un proceso de muestreo donde puede producirse el fenómeno citado. El lector recordará los *westerns* donde las ruedas de las diligencias semejaban girar en sentido contrario de lo normal. El efecto se debe a que la frecuencia de muestreo no es suficientemente alta comparada con la velocidad de giro de las ruedas, que si giraban 350° entre cuadros, para el espectador semejaban un movimiento contrario de 10° .

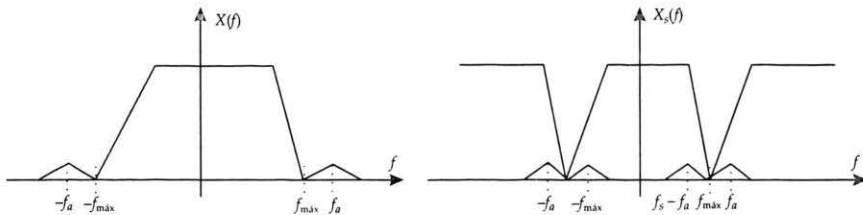


Figura III.12. a) Espectro de un mensaje no limitado en banda, b) espectro del mensaje muestreado

Como ejemplo podemos considerar una señal de $7/8$ Hz, la cual se muestrea a una frecuencia de 1 Hz. El análisis espectral nos indica que aparecerá una componente en $1/8$ Hz previamente inexistente. Es interesante observar el fenómeno en el dominio temporal, donde puede verse (figura III.13) que los valores de las muestras de la señal de $7/8$ Hz coinciden con los de la señal de $1/8$, y además que es la única señal de menor frecuencia que puede pasar por esos puntos.

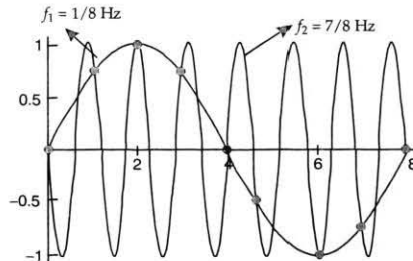


Figura III.13. Solapamiento de una señal de $7/8$ Hz en $1/8$ Hz

En la práctica siempre habrá un cierto grado de distorsión por aliasing, a pesar de haber prefiltrado la señal antes de muestrearla, lo que siempre es útil evaluar.

Sea $f(t)$ la señal de entrada que se limita en banda mediante un filtro

$$H(w) = \begin{cases} 1, & \text{para } |w| < W_f \\ \left(W_f/w\right)^n, & \text{para } |w| > W_f \end{cases} \quad (\text{III.20})$$

donde n es un número entero, característico del filtro, que nos da la pendiente de atenuación. La frecuencia de muestreo está dada por

$$w_s = 2KW_f \quad \text{donde } K > 1 \quad (\text{III.21})$$

Puesto que la pendiente de atenuación del filtro pasa bajas es finita, siempre existirá distorsión por adyacencia si $f(t)$ presenta componentes mayores que $w_s - W_f$. La potencia de señal en la banda de 0 a W_f es

$$P = \frac{1}{2\pi} \int_0^{W_f} |F(w)|^2 dw \quad (\text{III.22})$$

donde se ha considerado sólo el lado positivo del espectro.

La potencia de las componentes que producen distorsión por solapamiento adyacente está dada por

$$P_{\text{dist.}} = \frac{1}{2\pi} \int_{w_s - W_f}^{w_s} |F(w)|^2 \left(W_f/w\right)^{2n} dw \quad (\text{III.23})$$

donde por razones prácticas se supuso que la gran contribución a la distorsión adyacente la produce el espectro centrado en w_s , pudiéndose despreciar las de orden superior.

Por lo tanto, y teniendo en cuenta la ecuación (III.21), la relación señal a distorsión por adyacencia es

$$(S/N_{DA})_{n,K} = 10 \log_{10} \frac{\int_0^{W_f} |F(w)|^2 dw}{\int_{(2K-1)W_f}^{W_f} |F(w)|^2 \left(W_f/w\right)^{2n} dw} \quad (\text{III.24})$$

donde se observa que para una señal dada la relación señal-distorsión por adyacencia depende de K , de la frecuencia de muestreo y de n el orden del filtro.

Ejemplo III.1

Para ilustrar el proceso de muestreo y reconstrucción consideremos una señal senoidal simple $f(t) = 10 \cos(2\pi \cdot 20 \cdot t)$, que se muestrea a una tasa de 28 y 56 muestras por segundo.

En este caso la frecuencia máxima, $f_{\text{máx}}$, coincide con la frecuencia de la senoide, es decir 20 Hz,

por lo que debería muestrearse a $f_s \geq 2 f_{\max} = 40$ Hz. Consecuentemente la tasa de 28 Hz está por debajo de la tasa de Nyquist, produciéndose un fenómeno de submuestreo. Desde el punto de vista de la frecuencia el espectro de $f(t)$ es

$$F(w) = 5 \cdot \delta(w - 2\pi 20) + 5 \cdot \delta(w + 2\pi 20)$$

con lo que el espectro de la señal muestreada es:

$$F_s(w) = \frac{5}{T_s} \cdot \sum_{n=-\infty}^{\infty} [\delta(w - 2\pi 20 - n\omega_s) + \delta(w + 2\pi 20 - n\omega_s)]$$

En la figura III.14a se indica el espectro de la señal $f(t)$ mientras que en la III.14b y III.14c se representan los espectros de la señal muestreada a las dos diferentes frecuencias. Se muestra además, en línea punteada, la respuesta del filtro de reconstrucción, pasa bajas ideal, cuya frecuencia de corte se ha tomado en $0.5 \cdot f_s$. Resulta evidente que en el caso de submuestreo recuperamos una señal de frecuencia incorrecta igual a $f_s - f_a = 28 - 20 = 8$ Hz. En la figura III.14c se observa la correcta reconstrucción de $f(t)$.

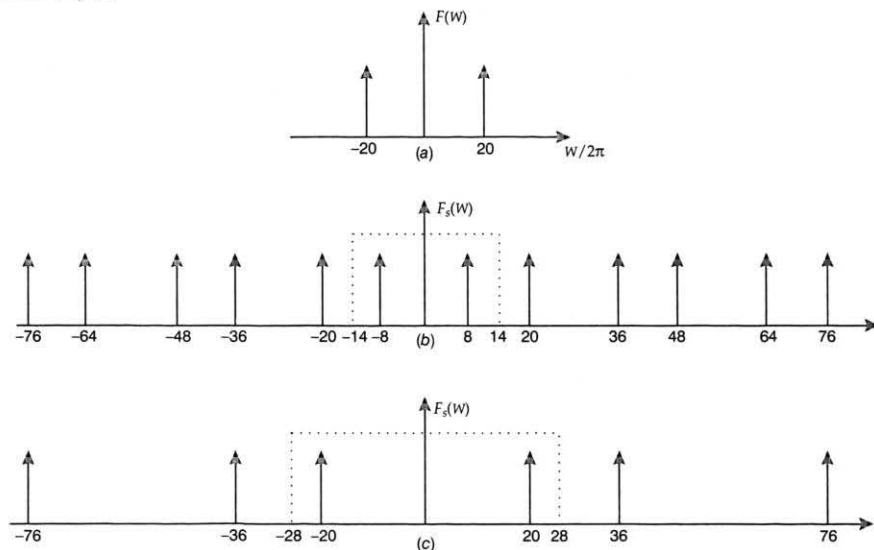


Figura III.14. a) Señal de entrada, b) señal submuestreada a $f_s = 28$ Hz, c) señal muestreada a $f_s = 56$ Hz

Ejemplo III.2

Consideremos ahora que $f(t)$ es una señal aperiódica, con lo que su espectro $F(w)$ será un continuo y, suponemos, limitado a $W_f = 20$ Hz (figura III.15a). Nuevamente tomaremos como frecuencias de muestreo 28 Hz y 56 Hz.

Para el caso del muestreo ideal, el espectro de la señal muestreada está dado por la ecuación (III.19), es decir que tendremos el espectro de la señal original multiplicado por f_s y luego reproducido en torno a todos los armónicos de la frecuencia de muestreo.

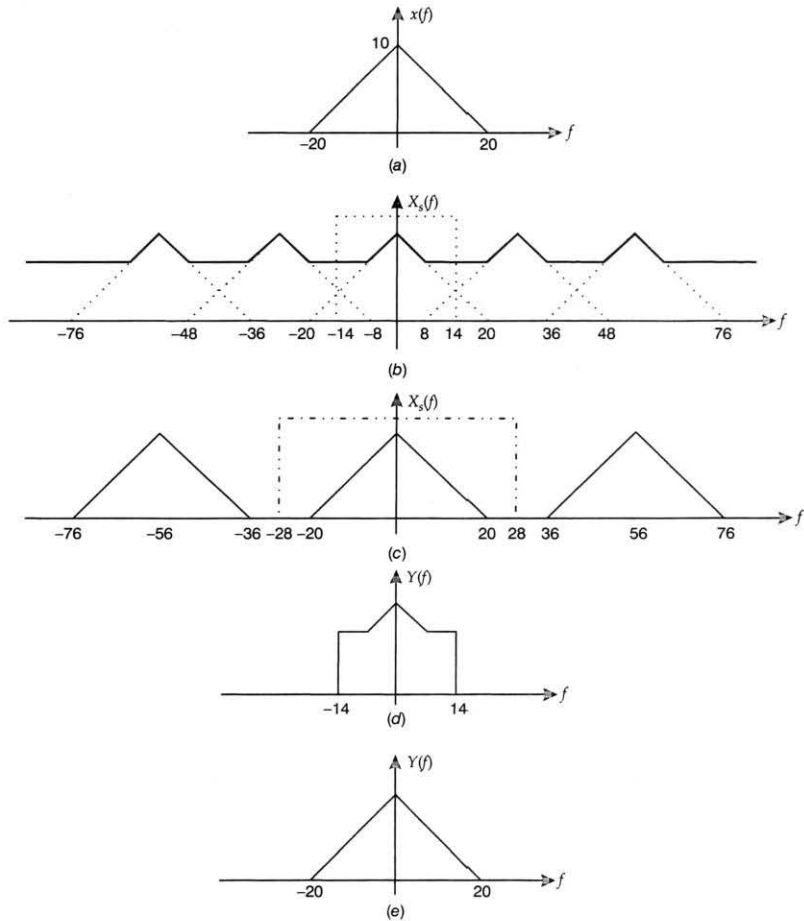


Figura III.15. a) Señal de entrada; b) señal submuestreada a $f_s = 28$ Hz; c) señal muestreada a $f_s = 56$ Hz; d) señal de salida con submuestreo; e) señal de salida

En las figuras III.15b y III.15c se indican los espectros que resultan para ambas frecuencias, pudiendo observarse en el caso $f_s = 28$ Hz el solapamiento de los espectros. Se ha indicado además la función de transferencia del filtro pasa bajas ideal usado para la reconstrucción, con frecuencia de corte $0.5 \cdot f_s$ y ganancia T_s . En las figuras III.15d y III.15e se representan los espectros de salida para ambos casos, observándose en el caso de submuestreo la discrepancia con el espectro original.

Ejemplo III.3

Suponga, a fin de fijar valores referenciales, un caso muy desfavorable donde el espectro de la señal de entrada $F(w)$ es plano desde cero a $f_s/2$ (ruido blanco limitado en banda), siendo $W_f \leq w_s/2$; es decir, que de 0 a $f_s/2$ está la señal nominal y de $f_s/2$ a f_s están las componentes indeseables. En la figura III.16 se ilustra esta situación.

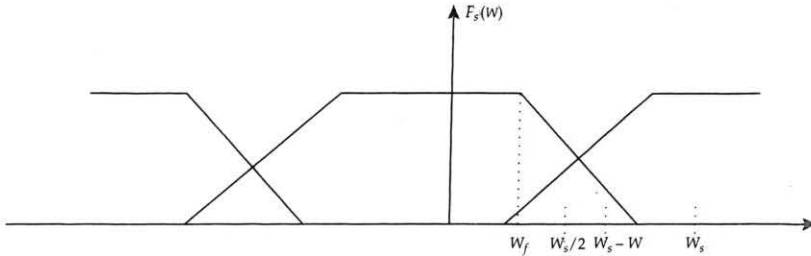


Figura III.16. Espectro de ruido blanco filtrado y muestreado ($w_s = 2KW_f$)

Calculemos la relación señal-distorsión adyacente para $K = 1; 1.5$ y 2 y $n = 1; 2; 3$ y 4 . Aplicando la ecuación (III.24) para

$$F(w) = \begin{cases} 1 & 0 \leq w \leq w_s \\ 0 & \text{para el resto} \end{cases}$$

se obtiene

$$(S/N)_{n,K} = 10 \cdot \log_{10} \left\{ \frac{1}{2n-1} \left[\frac{1}{(2K-1)^{2n-1}} - \frac{1}{(2K)^{2n-1}} \right] \right\} \quad (\text{III.25})$$

la cual para los valores dados nos da el siguiente cuadro

CUADRO III.1. Relación señal-distorsión por adyacencia

K	n	S/N	K	n	S/N	K	n	S/N
1	1	3.01	1.5	1	7.78	2	1	10.79
1	2	5.35	1.5	2	15.33	2	2	21.46
1	3	7.13	1.5	3	22.65	2	3	32.02
1	4	8.49	1.5	4	29.79	2	4	42.47

Los resultados del cuadro aparecen en la figura III.17

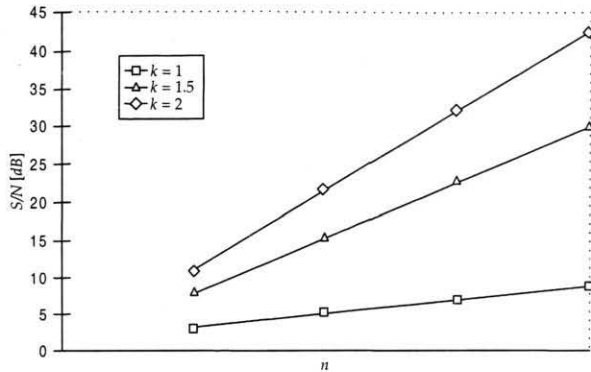


Figura III.17. Variación de la relación señal-distorsión por adyacencia

Este ejemplo muestra claramente que el filtrado previo controla la distorsión por solapamiento adyacente. Obsérvese que a mayor filtrado previo menor distorsión, y que aumentando la frecuencia de muestreo se inserta una *banda de guarda* entre $F(w)$ y las bandas adyacentes $F(w - w_s)$; $F(w - 2w_s)$, etcétera.

Resulta claro, además, que muestreando cerca de la frecuencia de Nyquist se obtiene muy poco beneficio aumentando el orden del filtro; para mayores valores de K , en cambio, la mejora obtenida con filtros más abruptos es sustancial.

Resumiendo, en el proceso de muestreo es fundamental limitar en banda la señal de entrada para minimizar la distorsión por solapamiento adyacente. Este filtrado previo incluso es recomendable aun cuando la señal a muestrear tenga escaso o nulo contenido armónico más allá de $w = w_s/2$. La razón es que a menudo las señales provenientes de sistemas de conmutación electromecánica, por ejemplo, pueden generar ruido en un amplio espectro, que al muestrearse produciría aliasing.

Ejemplo III.4

Dada la señal $x(t) = e^{-\alpha t} \cdot u(t)$, que no es limitada en banda, se desea determinar la mínima frecuencia de muestreo (expresada en anchos de banda de -3 dB de $x(t)$) de forma tal que la máxima componente introducida por aliasing esté a -10 dB de la máxima componente de la señal.

Puesto que $x(t) = e^{-\alpha t} \cdot u(t)$, su transformada de Fourier es

$$X(w) = \frac{1}{\alpha + jw} \quad \text{siendo} \quad |X(w)|^2 = \frac{1}{\alpha^2 + w^2}$$

Resulta claro que el ancho de banda de 3 dB en Hz será $BW_3 = \alpha/2\pi$. El espectro de la señal muestreada $X_s(w)$ aparece en la figura III.18, el máximo valor de $X_s(w)$ ocurre en $w = 0$ y resulta

$$X_s(f)|_{\text{máx}} = \frac{f_s}{\alpha}$$

mientras que el máximo aliasing ocurre a $f = f_s/2$, con lo que la máxima componente de aliasing es

$$a_{\text{máx.}} = \frac{f_s}{\sqrt{\alpha^2 + (2\pi)^2 \cdot (f - f_s)^2}} = \frac{f_s}{\sqrt{\alpha^2 + \pi^2 \cdot f_s^2}}$$

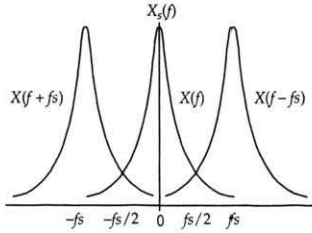


Figura III.18. Muestreo con solapamiento espectral

La condición del problema establece

$$\frac{X_s(f)|_{\text{máx.}}}{a_{\text{máx.}}} > 10 \quad \text{o sea} \quad \frac{\sqrt{\alpha^2 + \pi^2 f_s^2}}{\alpha} > 10$$

de donde resulta $f_s > 20 \cdot \frac{\alpha}{2\pi}$, es decir $f_s > 20 \cdot BW_3$.

III.5. MUESTREO NATURAL

El muestreo natural significa hacer una operación de multiplicación entre $f(t)$ y un tren de pulsos rectangulares, como se ve en la figura III.19a. En esta figura también se muestra la señal $f(t)$ y su espectro $F(w)$, así como el tren de pulsos en el dominio del tiempo y su respectivo espectro, y la señal muestreada obtenida del proceso de multiplicación tanto en el dominio del tiempo como en la frecuencia; analizando estas gráficas se observa que el espectro de la señal muestreada obtenido en este caso es muy similar al caso ideal, cuya principal diferencia es que sus componentes espectrales están moduladas por una función sinc; en este caso también se observa que es posible recuperar la señal original pasando las muestras a través de un filtro pasa bajas.

El análisis gráfico que aparece en la figura III.19 indica el proceso que debe realizarse para demostrar los resultados del muestreo natural. Partiremos de la serie de Fourier de un tren de pulsos de amplitud A y anchura τ y frecuencia igual a $w_s = 2\pi/T_s$, el cual está dado por

$$s(t) = \frac{A\tau}{T_s} \sum_{k=-\infty}^{\infty} \frac{\text{sen}(k w_s \tau/2)}{k w_s \tau/2} e^{j k w_s t} \quad (\text{III.26})$$

El espectro de $s(t)$ está dado por

$$s(w) = \frac{2\pi\tau A}{T_s} \sum_{k=-\infty}^{\infty} \frac{\text{sen}(k w_s \tau/2)}{k w_s \tau/2} \delta(w - k w_s) \quad (\text{III.27})$$

La señal muestreada $f_s(t)$ es el producto de $f(t)$ con $s(t)$

$$f_s(t) = s(t)f(t) \quad (\text{III.28})$$

El espectro $F_s(w)$ de $f_s(t)$ permite visualizar como puede recuperarse $f(t)$ a partir de sus muestras. Recordando que el producto en el tiempo de dos funciones tiene un espectro dado por $1/2\pi$ veces, la convolución de los respectivos espectros se obtiene

$$F_s(w) = \frac{1}{2\pi} F(w) * S(w)$$

$$F_s(w) = \frac{\tau A}{T_s} \sum_{k=-\infty}^{\infty} \frac{\text{sen}(kw_s \tau/2)}{kw_s \tau/2} \int_{-\infty}^{\infty} \delta(x - kw_s) F(w - x) dx \quad (\text{III.29})$$

$$F_s(w) = \frac{\tau A}{T_s} \sum_{k=-\infty}^{\infty} \frac{\text{sen}(kw_s \tau/2)}{kw_s \tau/2} F(w - kw_s)$$

Si $w_s \geq 2W_f$ el espectro de la señal muestreada contiene réplicas del espectro original, centradas a las frecuencias kw_s , las cuales no se traslapan, pero sí están escaladas por la función sinc.

Si $f_s(t)$ se pasa a través de un filtro pasa bajas con frecuencia de corte W_f , la señal original se puede recuperar sin distorsión, es decir en la sumatoria de III.29 sólo se toma el término $k=0$, entonces el espectro de la señal a la salida del filtro $S_0(w)$ es

$$S_0(w) = \frac{A\tau}{T_s} F(w) \quad (\text{III.30})$$

y la señal de salida en el dominio del tiempo es

$$s_0(t) = \frac{A\tau}{T_s} f(t) \quad (\text{III.31})$$

La salida es igual a la función original escalada por el factor $A\tau/T_s$, donde A = altura del pulso; τ = anchura del pulso y T_s = periodo de muestreo.

En el análisis anterior demostramos que el producto de un mensaje $f(t)$ y un tren de pulsos rectangulares de amplitud finita permite obtener un método de muestreo físicamente realizable. El espectro de la señal muestreada naturalmente $f_s(t)$ de $f(t)$ contiene réplicas sin distorsión del espectro del mensaje $F(w)$. El término central ($k=0$) es justo una versión escalada de $F(w)$ que permite la recuperación de $f(t)$ utilizando un filtro pasa bajas.

La realización práctica del multiplicador analógico de la figura III.19 es relativamente simple, tomando en cuenta la característica de la señal de muestreo y suponiendo que tiene amplitud unitaria, el proceso de muestreo natural se puede visualizar más que como un proceso de multiplicación como un proceso de conmutación.

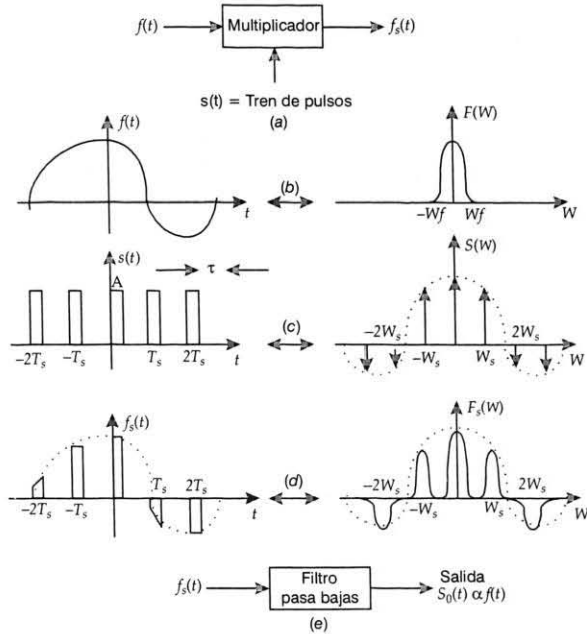


Figura III.19. a) Método de muestreo natural para la forma de onda de b) usando el tren de pulsos de c) para obtener la señal y espectro de d). La señal se recupera con el filtro pasa bajas de e)

III.6. MUESTREO Y RETENCIÓN

En este tipo de muestreo la amplitud de cada pulso es constante durante la duración del pulso, pero está determinada por la muestra instantánea de $f(t)$, tal como se ilustra en la figura III.20.

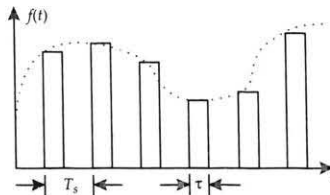


Figura III.20. Muestreo y retención de una señal

Suponiendo pulsos rectangulares ideales, la expresión para la señal muestreada es

$$f_s(t) = A \sum_{k=-\infty}^{\infty} f(kT_s) \text{Rect}\left(\frac{t - kT_s}{\tau}\right) \quad (\text{III.32})$$

donde la función Rect (x) está definida en el apéndice A.

Para encontrar la transformada de Fourier de la señal muestreada vamos a considerar que la señal de muestreo y retención puede obtenerse efectuando primero un muestreo ideal y después pasando estas muestras ideales por un circuito que retenga su valor durante cierto tiempo, como se observa en la figura III.21.

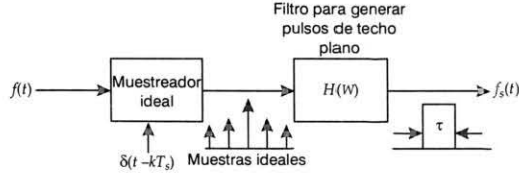


Figura III.21. Método de generación de muestreo y retención

La respuesta al impulso del filtro debe ser

$$h(t) = \text{Rect}\left(\frac{t}{\tau}\right) \leftrightarrow H(w) = \tau \cdot \frac{\text{sen}(w\tau/2)}{w\tau/2} \quad (\text{III.33})$$

el espectro de la señal muestreada está dado por:

$$F_s(w) = H(w) \cdot F(w) = \frac{\tau A}{T_s} \sum_{k=-\infty}^{\infty} \frac{\text{sen}(w\tau/2)}{w\tau/2} \cdot F(w - kw_s) \quad (\text{III.34})$$

El espectro antes determinado parece en primera instancia ser muy similar al de la ecuación (III.29) para muestreo natural; sin embargo, hay una importante diferencia, una operación de filtraje pasa bajas no proporciona una señal sin distorsión proporcional a $f(t)$; la señal correspondiente a la salida del filtro pasa bajas está dada por

$$S_0(w) = \frac{\tau A}{T_s} \left(\frac{\text{sen}(w\tau/2)}{w\tau/2} \right) F(w) \quad (\text{III.35})$$

que como claramente se ve no es proporcional a $F(w)$ como necesitamos; el factor $Q(w) = \text{sinc}(w\tau/2)$ representa distorsión que puede corregirse sumando un segundo filtro, denominado filtro igualador, el cual debe tener una función de transferencia $H_{eq} = 1/Q(w)$ para $w \leq W_f$, esto es

$$H_{eq}(w) = \begin{cases} \frac{w\tau/2}{\text{sen}(w\tau/2)} & \text{si } |w| \leq W_f \\ \text{arbitrario en otro caso} \end{cases} \quad (\text{III.36})$$

luego, para recuperar la señal original, aparte de pasar las muestras de la señal por un filtro pasa

bajas, la salida de éste hay que pasarla a través de un filtro igualador para eliminar la distorsión producida por el término sinc ($w\tau/2$); este proceso de recuperación se muestra en la figura III.22.

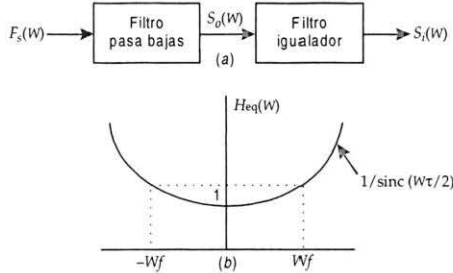


Figura III.22. a) Método de recuperación para muestreo y retención, el cual requiere un filtro de equalización con función de transferencia representada en b)

El espectro de salida equalizado es

$$S_i(w) = H_{eq}(w) \cdot S_0(w) = \frac{\tau A}{T_s} \cdot \left(\frac{\text{sen}(w\tau/2)}{w\tau/2} \right) \cdot \left(\frac{w\tau/2}{\text{sen}(w\tau/2)} \right) \cdot F(w) \quad (\text{III.37})$$

por lo tanto la señal de salida en el dominio del tiempo es

$$s_i(t) = \frac{A \tau}{T_s} f(t) \quad (\text{III.38})$$

La presencia de $Q(w)$ en la ecuación (III.35) puede entenderse como un término que distorsiona el espectro de $F(w)$ y en consecuencia el espectro de la señal recuperada; atenuándolo en las frecuencias altas resulta claro que cuanto más angosto sea el pulso más plano resultará su espectro en la banda $f < W_f$ y menor será la distorsión introducida. Este tipo de distorsión se denomina *distorsión por apertura*, pues se debe a que el impulso original (muestreo instantáneo) se ha abierto degenerando en un pulso de duración finita τ .

A fin de ejemplificar, suponga que $F(w)$ es plano y tiene un espectro extendido (W_f), el cual se representa en la figura III.23a (para $w > 0$ por simplicidad). En la figura III.23b aparece el espectro muestreado idealmente, mientras que en c se indica el del pulso rectangular. Finalmente en d tenemos el producto de ambos, tal como lo indica la ecuación (III.35).

Es evidente que las más altas frecuencias son las que resultan más atenuadas debido a la distorsión por apertura. Como ya hemos expresado, un método inmediato para disminuir la distorsión por apertura es reduciendo la duración del pulso haciendo $\tau \ll T_s$, que implica:

- Como $\tau \ll T_s$, la función sinc que pondera el espectro de salida se extiende haciéndose más plana en las frecuencias de interés.
- La relación τ/T_s , es decir el ciclo de trabajo, atenúa el espectro de salida.

Dicha atenuación puede compensarse insertando un amplificador, el cual también puede

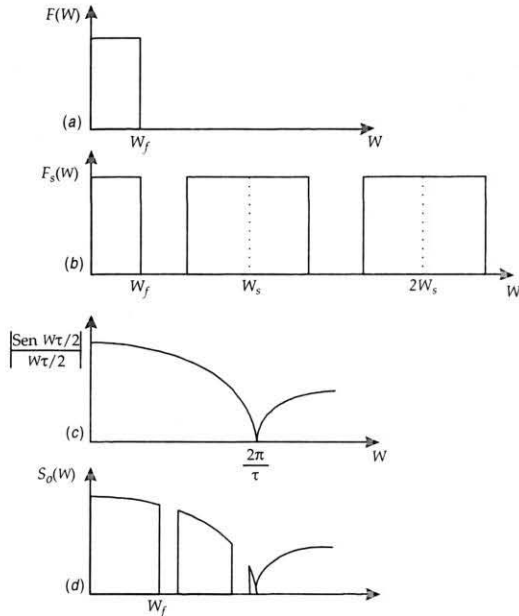


Figura III.23. Espectros de un sistema de muestreo y retención: a) de entrada, b) de la señal muestreada, c) del pulso, d) de la señal de salida

ecualizar la distorsión por apertura. En el cuadro III.2 se indica la ganancia en función de la frecuencia para distintos anchos de pulso.

CUADRO III.2.

$\tau/\text{Frecuencia}$	Ganancia en dB			
	0	$f_s/4$	$f_s/3$	$f_s/2$
T_s	0	0.91	1.65	3.92
$0.75 T_s$	2.5	3.0	3.41	4.61
$0.5 T_s$	6.02	6.24	6.42	6.93
$0.25 T_s$	12.04	12.10	12.14	12.26

Obsérvese que cuanto más angostos son los pulsos mayor debe ser la ganancia, pero más plana su respuesta. Así, por ejemplo, para $\tau = 0.25 T_s$ prácticamente no se requiere ecualización para compensar la distorsión por apertura.

III.7. MULTICANALIZACIÓN POR DIVISIÓN EN EL TIEMPO

Como mencionamos al inicio de este capítulo, uno de los grandes beneficios del muestreo es la posibilidad de realizar una técnica de multicanalización conocida como *Multicanalización por división en el tiempo* (TDM). La TDM permite, usando muestreo, que sea posible transmitir muchos mensajes sobre el mismo enlace de comunicación intercalando muestras de distintos mensajes.

La realización conceptual de la multicanalización en tiempo de N mensajes similares $f_n(t)$, $n = 1, 2, \dots, N$, está ilustrada en la figura III.24. Las señales muestreadas (trenes de pulsos) para los mensajes uno y dos aparecen en *a* y *b*. El tren de pulsos de *b* se retrasa ligeramente del tren de *a* para prevenir traslapamiento. Los otros mensajes se tratan en forma similar. Cuando el total de N trenes se combinan (multicanalizan) se obtiene la forma de onda de *c*. El tiempo asignado a cada una de las muestras se denomina *ranura de tiempo* (*time slot*). El intervalo de tiempo en el cual todos los mensajes se muestrean una vez se conoce como *trama* (*frame*). La porción de la ranura de tiempo no usada por la muestra se llama *tiempo de guarda* (*guard time*). En la figura III.24 todas las ranuras de tiempo están ocupadas por muestras de los mensajes. En un sistema práctico algunas de estas ranuras de tiempo pueden asignarse para otras funciones (por ejemplo señalización, monitoreo y sincronización).

III.7.1. Ancho de banda de TDM

Para calcular el ancho de banda de una señal TDM podemos suponer, sin pérdida de generalidad, que el tren de pulsos de la señal multicanalizada tiene una amplitud constante (como se ve en la figura III.25); estrictamente el ancho de banda de un tren de pulsos rectangulares de anchura τ es infinito, pero en una primera aproximación podemos suponer que el espectro extendido del tren de pulsos está dado por $W_f = 2\pi/\tau$; ($f_{\max} = 1/\tau$). Véase ejercicio uno al final del capítulo.

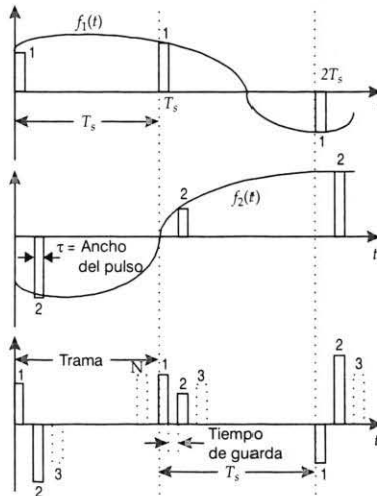


Figura III.24. Mensajes muestreados multicanalizados por división de tiempo. Trenes de pulsos de: a) mensaje 1, b) mensaje 2 y, c) el tren multiplexado

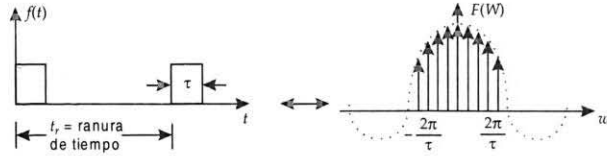


Figura III.25. a) Señal TDM simplificada, b) espectro de la señal TDM

Entonces, el ancho de banda de la señal TDM es

$$B_{TDM} = \frac{1}{\tau} \text{ (Hertz)} \quad (\text{III.39})$$

Un caso especial sucede cuando el tiempo de guarda es cero, como se ilustra en la figura III.26.

$$\Rightarrow B_{TDM} = \frac{1}{\tau} \text{ pero } \tau = \frac{T_s}{N}$$

con lo que se deduce

$$B_{TDM} = \frac{N}{T_s} = N \cdot f_s \text{ (Hertz)} \quad (\text{III.40})$$

es el ancho de banda para N señales multicanalizadas, en el caso de que el tiempo de guarda sea cero.

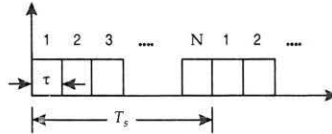


Figura III.26. Señal TDM sin tiempo de guarda

Ejemplo III.5

Supongamos que queremos multiplexar $N = 50$ señales similares y que cada señal está limitada a 3.3 KHz, las cuales son muestreadas a una frecuencia de 8 KHz. Se desea que el tiempo de guarda sea igual a la duración del pulso τ .

a) Encuentre el valor respectivo de τ y el ancho de banda necesario para transmitir esas señales.

$$t_r = \frac{T_s}{N} \quad \text{Además } t_r = 2 \tau$$

con lo cual se obtiene que

$$\Rightarrow \tau = \frac{T_s}{2N} = \frac{\frac{1}{f_s}}{2(50)} = \frac{1}{100f_s} = \frac{1}{8 \times 10^3 \times 10^2} = \frac{1}{8} \times 10^{-5} = 1.25 \mu \text{ seg}$$

$$B_{\text{TDM}} = \frac{1}{1.25 \times 10^{-6}} = 0.8 \times 10^6 = 800 \text{ KHz}$$

b) Si el tiempo de guarda es cero, ¿cuál es el ancho de banda requerido?

$$B_{\text{TDM}} = \frac{N}{T_s} = \frac{N}{\frac{1}{f_s}} = N \cdot f_s = 50 (8 \text{ KHz}) = 400 \text{ KHz}$$

III.7.2. Sincronización

Para mantener las posiciones apropiadas de las muestras en el multicanalizador es necesario sincronizarlo al proceso de muestreo. Debido a que las operaciones de muestreo usualmente son electrónicas, se utiliza un tren de pulsos como señal de reloj, que sirve de referencia para todos los muestreadores. En el receptor existe una señal de reloj que debe sincronizarse con la del transmisor. Una técnica de sincronización de reloj consiste en utilizar el lazo de sujeción de fase (PLL), para que sincronice la señal de reloj local con una derivada de la forma de onda recibida.

La sincronización de reloj no garantiza una sincronización apropiada de la secuencia de muestras. La alineación de las ranuras de tiempo requiere de sincronía de trama. Una técnica sencilla consiste en utilizar una o más ranuras de tiempo por trama para sincronización. Como ejemplo se puede colocar un pulso especial en la primera ranura, que sea más grande que la mayor amplitud esperada del mensaje; el inicio de trama fácilmente puede identificarse usando un circuito de comparación apropiado.

III.8. INTERFERENCIA ENTRE PULSOS (ISI)

Al analizar los sistemas de comunicaciones digitales, muchas veces ignoramos la forma de los pulsos que se usan para transmitir la información. En muchos ejemplos simplemente mostramos que los pulsos tienen una forma rectangular, pero en la práctica, donde estos pulsos pueden modular una portadora para transmisión sobre relativamente grandes distancias, debe considerarse su conformación; particularmente esto ocurre donde hay restricciones sobre el ancho de banda del canal.

En una primera aproximación por lo regular se dice que el ancho de banda está determinado por el recíproco de la ranura de tiempo o intervalo en el cual el pulso está restringido a situarse. Entonces, si 10 señales son muestreadas y multiplexadas en tiempo cada 125 μs , cada muestra de la señal está restringida a situarse dentro de una ranura de tiempo de 12.5 μs . El ancho de banda requerido para transmitir la señal TDM es entonces $1/12.5 \mu\text{s}$, o 80 KHz. Sin embargo, recordemos que la relación anterior sólo es una aproximación para calcular el ancho de banda de un tren de pulsos, y que estrictamente su ancho de banda es infinito. Sin embargo, al tratar la conformación de pulsos debemos ser más precisos al determinar el ancho de banda. Para analizar más detalladamente el problema, supongamos que deseamos transmitir un tren de pulsos a través de un canal con un ancho de banda finito de B Hz, como se muestra en la figura III.27.

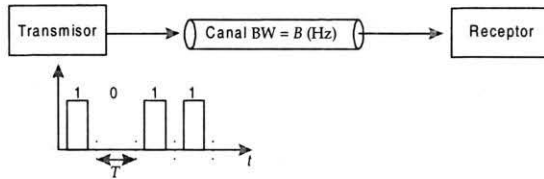


Figura III.27. Sistema básico de comunicación con un canal cuyo ancho de banda es finito

Como ya hemos dicho, el ancho de banda estrictamente necesario para transmitir el tren de pulsos es infinito; sin embargo, el canal de transmisión se comporta como un filtro pasa bajas que al atenuar las componentes de alta frecuencia provoca que los pulsos se dispersen y el traslapamiento de los pulsos entre ranuras de tiempo adjuntas puede ser muy fuerte, lo que podría manifestarse en una detección errónea en el receptor. Si tomamos como ejemplo el caso de la figura III.27, el 0 transmitido puede aparecer como un 1 si la suma de las colas de los pulsos adyacentes provocan un valor alto (en la práctica hay más contribuciones debidas a las colas de más pulsos adyacentes y no sólo dos como se muestra en la figura III.28). Este fenómeno de traslapamiento y la dificultad resultante en el proceso de decisión del receptor se conoce como *interferencia entre pulsos (ISI)*.

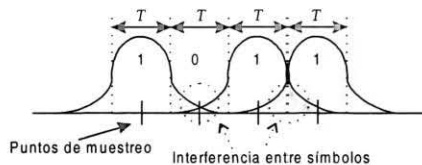


Figura III.28. Interferencia entre símbolos en transmisión digital

Una solución a este problema es reducir la anchura de los pulsos, pero como ya dijimos, eso provoca que aumente el ancho de banda de transmisión; en ese caso la solución está en diseñar una conformación de pulso adecuada y un filtro que minimice o elimine esta interferencia entre símbolos, con un ancho de banda de transmisión tan pequeño como sea posible.

Una conformación de pulso que produce cero interferencia entre símbolos justamente es el pulso $\text{sinc}(x)/x$, introducido en la sección III.3.1; como la respuesta al impulso de un filtro ideal pasa bajas, específicamente si el filtro tiene un espectro de amplitud constante hasta B (Hz), y es cero en todas las demás frecuencias, la respuesta al impulso es justamente $(\text{sen } 2\pi Bt)/2\pi Bt$. Este filtro y su respuesta al impulso se ven en la figura III.29. Cabe hacer notar que los cruces por cero de estos puntos están igualmente espaciados y ocurren en múltiplos de $1/2B$ segundos. Si se selecciona $1/2B$ como el intervalo de muestreo T , es evidente que pulsos conformados en la misma forma y con amplitud arbitraria estarán espaciados por intervalos $T = 1/2B$ segundos, no se interferirán entre sí, aunque en este caso hay traslapamiento entre pulsos, ya que en el instante de muestreo el pulso de interés es máximo y los demás son cero la interferencia es nula, esto se ilustra en la figura III.30.

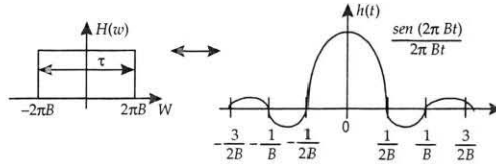


Figura III.29. Forma de pulso con cero interferencia entre símbolos

Los puntos de cruce están dados por:

$$2\pi Bt = n\pi \Rightarrow t = \frac{n}{2B}, n = 1, 2, 3, 4 \dots$$

A partir de la figura III.30, si n = número de pulsos, entonces en el caso ideal (filtro ideal) se pueden transmitir sin interferencia hasta $2B$ pulsos por segundo. Dicho de otra manera, si tenemos un canal de transmisión cuyo ancho de banda es B hertz, a través de él se pueden transmitir hasta $2B$ pulsos por segundo. Esto se obtiene a partir de la siguiente relación.

$$\frac{n}{2B} = 1 \text{ seg} \Rightarrow n = 2B \text{ pulsos/seg} \quad (\text{III.41})$$

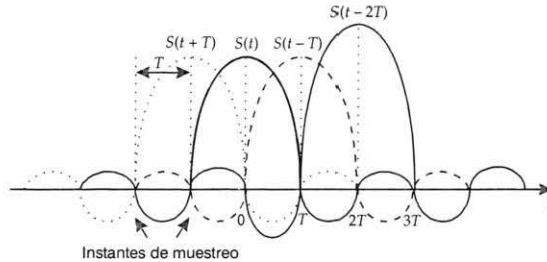


Figura III.30. Secuencia de pulsos con cero interferencia entre pulsos

Los anteriores resultados fueron establecidos por Nyquist como un teorema, el cual establece:

Si r símbolos independientes por segundo se transmiten a través de un canal ideal pasa bajas de ancho de banda B (Hz), la máxima velocidad de $r = 2B$ símbolos por segundo puede ser alcanzada sin interferencia entre símbolos si el pulso se conforma como un pulso sinc de la forma $\text{sinc}(\pi r t)$.

Ejemplo III.6

Calcule la máxima velocidad que puede transmitirse a través de un canal telefónico.

Si B = Ancho de banda telefónico = 3 100 Hz, entonces

$$r = 2B = 2 \times 3.1 \text{ KHz} = 6.1 \text{ K símbolos/seg}$$

Este resultado puede parecer paradójico con la velocidad actual que alcanzan los módems de alta velocidad, aunque la contradicción se resuelve si recordamos lo visto en el capítulo II, donde se explicó que un símbolo puede llevar más de un bit de información; en particular los módems de alta velocidad usan esquemas de modulación multinivel, donde cada símbolo transmite más de un bit de información.

Existen diversas dificultades prácticas para realizar la conformación de pulso anteriormente descrita:

- La conformación de pulso analizada implica que las características totales entre el transmisor y el punto de decisión del receptor correspondan a las de un filtro pasa bajas ideal, físicamente no realizable, y muy difícil de aproximar en la práctica por la atenuación infinita que presenta a la frecuencia de corte.
- Si este pulso en particular fuese realizable, requeriría de una sincronización muy precisa entre transmisor y receptor, ya que si ésta varía, desaparece la condición de cero interferencia entre símbolos. De hecho, bajo ciertas secuencias de pulsos, las colas de todos los pulsos adyacentes pueden sumarse y dan como resultado una serie divergente, causando posibles errores.

Sin embargo, a partir de este pulso ideal es posible derivar otras conformaciones de pulso con cero ISI que superan las dos dificultades mencionadas, mucho más simples de obtener en forma práctica, y en las que se pueden reducir los efectos de sincronización. Las condiciones para obtener estos pulsos fueron establecidas por Nyquist en su teorema conocido como *teorema de simetría vestigial*, que establece:

Si $P_r(f)$ (espectro del pulso) satisface

$$\sum_{k=-\infty}^{\infty} P_r(f + k/T_b) = K \text{ para } |f| < \frac{1}{2T_b} \quad (\text{III.42})$$

con K una constante arbitraria y T_b es la duración del pulso, entonces $p_r(t)$ no produce interferencia entre símbolos.

No demostraremos aquí este teorema (véase por ejemplo Shanmugan [1]). A primera vista la ecuación (III.42), puede parecer difícil de interpretar; note que el criterio de la ecuación (III.42) no especifica únicamente la forma espectral de $P_r(f)$. Entonces ¿cuál es su significado?

Trataremos de aclarar la interpretación de la ecuación (III.42) si seleccionamos para propósitos de ilustración una función específica $P_r(f)$ que satisfaga el criterio. Consideremos como ejemplo la forma espectral que aparece en la figura III.31a. Es importante aclarar que ésta no es una función práctica para $P_r(f)$, sino simplemente ha sido seleccionada para explicar el significado de la ecuación que nos ocupa.

Expandiendo la ecuación y aplicándola al pulso de la figura III.31a obtenemos

$$\dots P_r(f - 1/T_b) + P_r(f) + P_r(f + 1/T_b) + \dots = T_b \text{ para } |f| < 1/2 T_b$$

aparentemente sólo los tres términos mostrados contribuyen a la sumatoria en la región $|f| < 1/2 T_b$. En las figuras III.31b y III.31c se muestran las otras dos funciones desplazadas que deben considerarse.

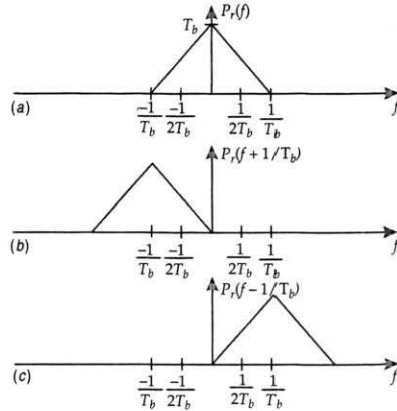


Figura III.31. Espectro de pulso que ilustra el criterio de simetría vestigial de Nyquist

Es claro que la suma es igual a T_b en el intervalo $|f| < 1/2 T_b$, de tal manera que la ecuación (III.42) es satisfecha.

¿Cómo podemos describir la clase de pulsos espectrales $P_r(f)$ que satisfacen la ecuación (III.42)? En resumen, el pulso básico debe tener una transformada de Fourier para que la parte de frecuencia positiva esté 6 dB por debajo a la frecuencia de $f = 1/2 T_b$, y la cual tenga simetría impar alrededor de esta frecuencia (de aquí se deriva el termino *simetría vestigial*), tal como se ilustra en la figura III.32.

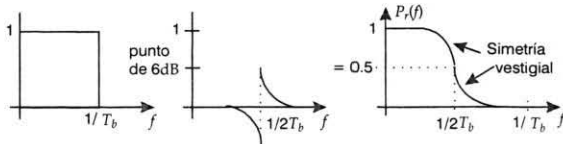


Figura III.32. Construcción de simetría vestigial para cero ISI

De hecho hay un número infinito de espectros que satisfacen este requerimiento; en las próximas secciones examinaremos una clase de pulsos que resultan de interés práctico.

III.8.1. Espectro de coseno elevado

Para obtener esta clase particular de espectro comenzaremos a partir del filtro ideal y modificaremos su característica alrededor de su frecuencia de corte para obtener un decaimiento más gradual, con la intención de encontrar un filtro físicamente realizable. En particular si la nueva característica de frecuencia es diseñada para tener simetría impar alrededor de la frecuencia de corte del filtro ideal (como lo establece el teorema de Nyquist), la respuesta al impulso resultante conserva la propiedad de tener cruces por cero a intervalos de tiempo uniformemente espaciados. Un ejemplo de este espectro que algunas veces se utiliza en la práctica es el *espectro de coseno elevado*. Este

espectro en particular y el espectro pasa bajas ideal original aparecen en la figura III.33. Para evitar confusiones con el ancho de banda de transmisión B , indicamos la frecuencia de corte del filtro pasa bajas ideal como w_c . El espectro de coseno elevado está dado por

$$H(\omega) = \begin{cases} \frac{1}{2} \left(1 + \cos \frac{\pi \omega}{2\omega_c} \right) & \text{si } |\omega| \leq 2\omega_c \\ 0 & \text{en otro caso} \end{cases} \quad (\text{III.43})$$

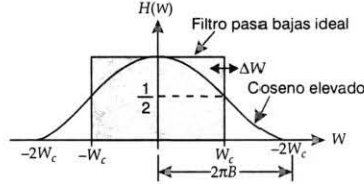


Figura III.33. Espectro de coseno elevado

El ancho de banda en este caso es $2\pi B = 2\omega_c$. Si ahora lo analizamos cuando hay una variación alrededor de ω_c , es decir $\omega = \omega_c + \Delta\omega$, podemos escribir el espectro como

$$H(\omega) = \frac{1}{2} \left[1 + \cos \frac{\pi (\omega_c + \Delta\omega)}{2\omega_c} \right] = \frac{1}{2} \left[1 + \cos \frac{\pi}{2} \left(1 + \frac{\Delta\omega}{\omega_c} \right) \right]$$

$$H(\omega) = \frac{1}{2} \left(1 - \sin \frac{\pi \Delta\omega}{2\omega_c} \right)$$

entonces el espectro cosenoidal desplaza la simetría impar alrededor de ω_c . La respuesta al impulso de un filtro con estas características está dada por:

$$h(t) = \frac{\omega_c}{\pi} \frac{\sin \omega_c \cdot t}{\omega_c \cdot t} \frac{\cos \omega_c \cdot t}{1 - \left(\frac{2\omega_c \cdot t}{\pi} \right)^2} \quad (\text{III.44})$$

Esta respuesta contiene el término $(\sin x)/x$ del filtro ideal, multiplicada por un factor adicional que decrece con el tiempo. El término $(\sin x)/x$ asegura que no hay interferencia entre símbolos; el factor adicional únicamente reduce la amplitud de las colas de los pulsos por un factor considerablemente menor que el término $(\sin x)/x$, lo que produce que este tipo de pulso sea menos sensible a la sincronización.

Haciendo el intervalo de muestreo $T = 1/2 f_c = \pi/\omega_c$, tal que los ceros de los pulsos ocurran a intervalos de T segundos, como en el caso previo del muestreo ideal, obtenemos que el ancho de banda de transmisión requerido es $B_{\cos} = 2 f_c = 1/T$. Como ejemplo, consideremos una señal analógica de 3.3 KHz muestreada a una razón de 8 KHz. Si esta señal fuese transmitida por TDM, y si se utilizase un filtro conformador ideal, el ancho de banda requerido será de $B_{\text{ideal}} = 1/2 T = 4$ KHz. Sin embargo, si utilizásemos un espectro de coseno elevado, el ancho de banda requerido sería de

$B_{\cos} = 1/T = 8 \text{ KHz}$. Si 10 señales fuesen multiplexadas en tiempo, el ancho de banda se incrementa por un factor de 10. Entonces la conformación de pulso ideal $(\text{sen } x)/x$ teóricamente permite la transmisión a un ancho de banda muy cercano al de la señal analógica original (el ancho de banda es igual al de la señal analógica si se utiliza una frecuencia de muestreo igual a la de Nyquist), por lo que el uso de pulsos conformados más reales provoca un incremento en el ancho de banda requerido. El espectro de coseno elevado duplica el ancho de banda requerido respecto al caso ideal. En la siguiente sección se analizarán otra clase de pulsos que poseen simetría vestigial pero que requieren menor ancho de banda.

III.8.2. Espectro de seno elevado

Otro tipo de espectro que tiene simetría impar y cero interferencia entre pulsos es el denominado *espectro de seno elevado*, que ilustramos en la figura III.34; su espectro de amplitud está especificado por:

$$|H(\Delta\omega)| = \begin{cases} \frac{1}{2} \left(1 - \text{sen} \frac{\pi \Delta\omega}{2 \omega_x} \right) & \text{Si } |\Delta\omega| < \omega_x \\ 0 & \text{Si } |\Delta\omega| > \omega_x \\ 1 & -\omega_c < \Delta\omega < -\omega_x \end{cases} \quad (\text{III.45})$$

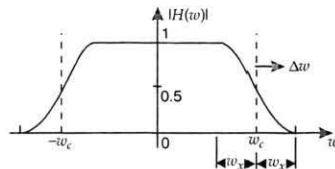


Figura III.34. Espectro de seno elevado

El parámetro de diseño w_x representa la frecuencia en radianes, por lo que el ancho de banda w_c se excede. La relación $\alpha = w_x/w_c$ se le conoce como factor de *roll-off*; pequeños valores de w_x/w_c permiten menores anchos de banda, pero requieren mayor complejidad en el diseño de los filtros. El caso de $w_x = 0$, o cero *roll-off*, justamente es el caso del filtro pasa bajas ideal que produce el pulso $(\text{sen } x)/x$ ya discutido. Similarmente un factor de *roll-off* de uno es decir $w_x/w_c = 1$, produce el espectro de coseno elevado. Entonces al variar el factor de *roll-off* se puede generar toda una clase de pulsos conformados; en la figura III.35 se muestra la forma del pulso conformado para varios valores de *roll-off*, es de notar que para $\alpha = 1$ (filtro de coseno elevado) existen exactamente la mitad de cruces por cero que para el caso de $\alpha = 0$ (filtro ideal), lo cual está relacionado con el número de pulsos por segundo que se pueden transmitir en cada caso. La respuesta al impulso del espectro de seno elevado está dada por:

$$h(t) = \frac{\omega_c}{\pi} \frac{\text{sen } \omega_c (t - t_0)}{\omega_c (t - t_0)} \frac{\cos \omega_x (t - t_0)}{1 - \left[\frac{2 \omega_x (t - t_0)}{\pi} \right]^2} \quad (\text{III.46})$$

En este caso existe un compromiso entre el ancho de banda y la velocidad de los pulsos, que depende del factor de *roll-off* seleccionado. Lo que esto indica es que si la transmisión de pulsos deseada es igual a r pulsos por segundo, el ancho de banda B requerido, en Hz, es

$$B = \frac{r}{2} \cdot \left(1 + \frac{f_x}{f_c} \right) = \frac{r}{2} (1 + \alpha) \quad (\text{III.47})$$

si conocemos B (el ancho de banda de transmisión), el número de pulsos por segundo que pueden ser transmitidos está dado por:

$$r = \frac{2B}{1 + \frac{f_x}{f_c}} = \frac{2B}{1 + \alpha} \Rightarrow \begin{cases} r = 2B & \text{caso ideal } (\alpha = 0) \\ r = B & \text{caso cosenoidal } (\alpha = 1) \end{cases} \quad (\text{III.48})$$

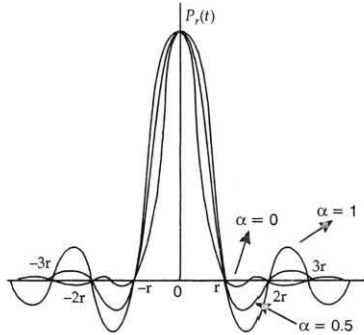


Figura III.35. Forma de pulso conformado con filtro de seno elevado para varios valores de α

III.8.3. Circuitos conformadores de pulsos

El diseño de los filtros conformadores de pulsos puede ser realizado por medio de alguno de los dos siguientes métodos.

Técnicas de filtrado analógico. Una red con la respuesta en frecuencia del espectro de seno elevado puede aproximarse por medio de técnicas convencionales de diseño de filtros. Por ejemplo, Feher [2] detalla la elaboración de un filtro elíptico de séptimo orden ecualizado en fase, el cual aproxima al pulso de seno elevado con $\alpha = 0.3$.

Técnicas de filtrado digital. Un método alternativo para generar pulsos conformados con las características descritas consiste en utilizar una estructura de filtro digital, como se ve en la figura III.36.

La secuencia de impulsos codificados $\{a_k\}$ se introduce a una línea de retraso constituida por $2N$ elementos de retraso, cada uno produce un retraso de Δ seg. Los coeficientes $b_{-N}, b_{-N+1} \dots b_0, b_1 \dots b_{N-1}, b_N$, son constantes multiplicativas que determinan la respuesta al impulso del filtro. Consideremos un pulso simple positivo de $+V$ entrando al filtro al tiempo $t = -N$ seguido por una larga cadena de ceros. La señal de reloj del filtro desplaza el pulso de entrada $+V$ a través

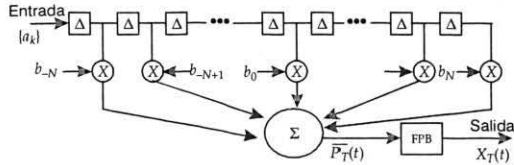


Figura III.36. Conformación de pulso utilizando un filtro transversal

de la línea de retardo a una velocidad varias veces más alta que la velocidad de los pulsos. En la práctica el número de elementos de memoria $2N$ puede ser típicamente de 20 a 30.

Luego, la salida del filtro transversal $\bar{p}_T(t)$ deberá tener la forma de una señal en escalera, como se muestra en la figura III.37. En esta figura, la curva sólida representa la respuesta al impulso deseada $p_T(t)$ del filtro. Los coeficientes del filtro $b_{-N}, b_{-N+1} \dots b_N$, corresponden a los valores de las muestras de $p_T(T)$. La salida $\bar{p}_T(t)$ está dada por:

$$\bar{p}_T = \sum_{j=-N}^N b_j \cdot s(t - \Delta j) \quad \text{donde} \quad s(t) = \begin{cases} 1 & -\Delta/2 < t < \Delta/2 \\ 0 & \text{en otro caso} \end{cases}$$

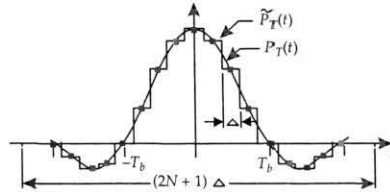


Figura III.37. Respuesta al impulso del filtro transversal

La salida $\bar{p}_T(t)$ puede suavizarse con un filtro pasa bajas simple. En el caso de una entrada consistente de pulso simple $+V$ seguida por ceros, la salida aproxima la conformación de pulso requerida para cero ISI.

En el ejemplo mostrado, el símbolo se desplaza por la línea de retardo a cuatro veces la velocidad de los datos, ya que se usa $4\Delta = T_b$. Para una entrada general, los sucesivos pulsos a_k se desplazarán en uno cada T_b seg, y el filtro hace la convolución de la secuencia de entrada con la respuesta al impulso del filtro. Esto es, la forma de onda de la salida $x_T(t)$ representa la superposición de pulsos sucesivos que resultan de los sucesivos pulsos de entrada en la secuencia $\{a_k\}$.

Ejemplo III.7

Consideremos una señal analógica limitada en banda a 3.2 KHz, la cual es muestreada a 8 KHz.

a) ¿Cuál es el ancho de banda de transmisión requerido si se utiliza un filtro ideal en el receptor?

$$B_i = f_c = \frac{r}{2} = \frac{f_s}{2} = 4 \text{ KHz}$$

b) ¿Cuál es el ancho de banda de transmisión requerido si se utiliza un filtro cosenoidal?

$$B_c = r = f_s = 8 \text{ KHz}$$

Ejemplo III.8

a) ¿Cuál es la máxima velocidad en pulsos/seg que se puede transmitir a través de la red telefónica si los pulsos se modulan como pulsos ideales, cosenoidales y senoidales $\alpha = 0.5$?

b) ¿Se puede transmitir a través de la red telefónica a una velocidad de 9 600 bits/seg?

Respuesta:

a) En este caso se debe hacer notar que el ancho de banda del canal es constante = 3 100 Hz.

$$r_{\text{ideal}} = 2B_{\text{canal}} = 6\,200 \text{ pulsos/seg}$$

$$r_{\text{cos}} = B_{\text{canal}} = 3\,100 \text{ pulsos/seg}$$

$$\left| r_{\text{sen}} = \frac{2B_{\text{canal}}}{(1 + \alpha)} \right|_{\alpha = 0.5} = \frac{6\,200}{1.5} = 4\,133 \text{ pulsos/seg}$$

Aquí podemos concluir, si $B_{\text{canal}} = \text{constante}$, la velocidad de transmisión la podemos aumentar disminuyendo el factor de *roll-off*, es decir acercándonos lo más posible al filtro ideal.

b) Sí, utilizando un sistema multinivel; por ejemplo, si cada pulso transporta 2 bits, y $\alpha = 0.25$.

$$r_{\text{sen}} = 4\,960 \text{ pulsos/seg} \cdot 2 \text{ bits/pulso} = 9\,920 \text{ bits/seg}$$

Ejemplo III.9

a) ¿Qué ancho de banda se necesitaría para transmitir a una velocidad de 9 600 bits/seg?

$$B_{\text{ideal}} = \frac{r}{2} = \frac{9\,600}{2} = 4\,800 \text{ Hz}$$

$$B_{\text{cos}} = r = 9\,600 \text{ Hz}$$

$$B_{\text{sen}} = \frac{r \cdot (2 + \alpha)}{2} = \frac{9\,600 \cdot (1 + 0.25)}{2} = 6\,000 \text{ Hz}$$

Si la velocidad es constante, el ancho de banda lo podemos disminuir si el sistema se aproxima al filtro ideal.

En general podemos decir que si N señales fuesen multicanalizadas en tiempo, el ancho de banda se incrementaría por un factor N

$$B_i = \frac{N \cdot r}{2} \text{ Hertz} \rightarrow \text{filtro ideal}$$

$$B_c = N \cdot r \text{ Hertz} \rightarrow \text{filtro cosenoidal} \quad (\text{III.49})$$

$$B_s = \frac{N \cdot r}{2} \cdot (1 + \alpha) \text{ Hertz} \rightarrow \text{filtro senoidal}$$

III.9. PROBLEMAS

- III.1. Calcule la energía en el lóbulo central de la siguiente función:

$$F(\omega) = \tau \operatorname{sinc}\left(\frac{\omega\tau}{2}\right)$$

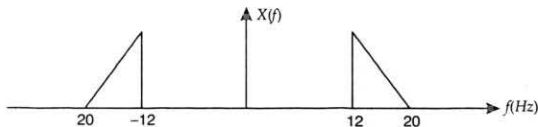
correspondiente a la transformada de Fourier por un pulso cuadrado de amplitud unitaria centrado en el origen y de ancho igual a τ .

- III.2. Determine la frecuencia de corte del filtro pasa bajas ideal utilizado para reconstruir la señal:

$$x(t) = 10 \cos[600\pi t] \cos^2[1600\pi t]$$

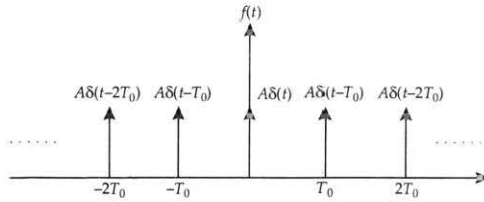
de la cual se toman 4000 muestras por segundo. Encuentre la frecuencia mínima de muestreo.

- III.3. La señal $x(t) = 6 + 4 \cos[10\pi t] + 4 \cos[14\pi t] + 2 \cos[20\pi t]$ se muestrea a una tasa de 30 muestras/seg. Grafique el espectro de la señal muestreada indicando todos los componentes para $f < 80$ Hz. Explique cómo se puede reconstruir $x(t)$ a partir de las muestras.
- III.4. Sea $x(t) = \cos[2\pi 100t] + \cos[2\pi 220t]$ que se muestrea a $f_s = 300$ Hz. Si la señal muestreada $x_s(t)$ pasa luego por un filtro pasa bajas ideal de frecuencia de corte 150 Hz, indique qué frecuencias componentes estarán presentes en la salida.
- III.5. La señal cuyo espectro aparece en la figura, idealmente se muestrea con $f_s = 20$ Hz. Se desea representar el espectro de la señal muestreada $x_s(t)$ para $f \leq 40$ Hz. ¿Puede recuperarse $x(t)$? Si es así, explique cómo lo haría.

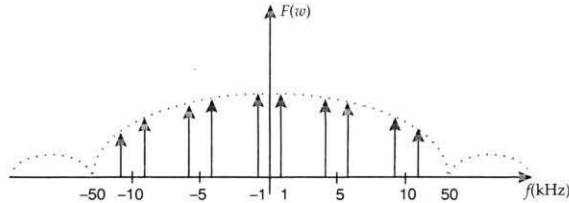


- III.6. Una señal $x_1(t)$ y otra señal $x_2(t)$ de frecuencia máxima 2 y 4 KHz respectivamente se multiplexan en tiempo usando una única frecuencia de muestreo. Calcule el intervalo de muestreo.
- III.7. Calcule la serie de Fourier trigonométrica y exponencial del tren de impulsos que aparece a continuación:

Muestreo



III.8. Una señal muestreada tiene un espectro como el que se observa en la siguiente figura:



- ¿Cuál es la señal que se está muestreando?
 - ¿Cuál es la frecuencia de muestreo?
 - Determine qué tipo de muestreo se está utilizando y la anchura de los pulsos de muestreo.
- III.9. Un canal de transmisión tiene un ancho de banda de 3.6 KHz.
- Calcule la velocidad de transmisión en bits/seg, si se transmiten pulsos binarios. Suponga que el sistema de transmisión se comporta como un filtro senoidal, con un factor de *roll-off* de $\alpha = 1, 0, 0.25$ y 0.5 .
 - Repita el inciso a) pero ahora considere que los pulsos transmitidos tienen 16 valores posibles.
- III.10. Un sistema transmite muestras ideales de la señal de entrada. La razón de muestreo es de 25 KHz y el receptor usa un filtro pasa bajas ideal con un ancho de banda de 10 KHz. La señal de entrada está limitada en banda a 13 KHz.
- ¿Puede reconstruirse esta señal sin distorsión?
 - Si no es posible, ¿qué cambios deben hacerse al sistema para recuperar la señal?
- III.11. Una señal $f_1(t) = 6 u(t) \exp(-3t)$ tiene una transformada de Fourier $F(\omega) = 6 [3 + j\omega]^{-1}$. Claramente se ve que $f_1(t)$ no está limitada en banda.
- Encuentre el ancho de banda de un filtro ideal tal que la señal filtrada denotada por $f(t)$ contenga 99% de la energía total.
 - ¿Cuál es la frecuencia mínima a la que se debe muestrear $f(t)$ sin producir traslapamiento (*aliasing*)?
- III.12. Una señal está dada por $f(t) = 3 \cos [2 \pi (10)^4 t]$.
- ¿Cuál es la frecuencia de Nyquist para esta señal?
 - Si esta señal se muestrea idealmente a una frecuencia de 60 kHz y se aplica a un filtro pasa bajas ideal cuyo ancho de banda es de 30.1 kHz, encuentre una ecuación para la señal $f_0(t)$, que representa la salida del filtro.
 - ¿Hay traslapamiento? Explique.
- III.13. Una señal $f(t) = 6 \cos [10^3 t]$ es muestreada naturalmente a una razón de $\omega_s = 2.5 (10^3)$ rad/seg

usando pulsos cuya anchura es de $60 \mu\text{s}$. La amplitud del tren de pulsos de muestreo es $A = 2$ volts.

a) Escriba una expresión para la respuesta del filtro ideal de recepción (pasa bajas) cuyo ancho de banda es de $\omega_c = 1010$ rad/seg.

b) Si ω_c es cambiada a 990 rad/seg, ¿cuál es la respuesta?

III.14. Una señal en banda base se reconstruye exactamente a partir de sus muestras, las cuales fueron tomadas a una frecuencia de 60 Hz. Si estas muestras fueron tomadas a tres veces la frecuencia de Nyquist, ¿cuál es el espectro extendido de la señal?

III.15. Una planta de potencia utiliza un multicanalizador de palabras para transmitir muchas señales de monitoreo a una consola de control principal. La línea tiene 500 ranuras de tiempo, de igual longitud, cada una transportando una palabra binaria de 4 bits naturales, las cuales son muestras a la velocidad de Nyquist de fuentes de mensajes similares de banda angosta, cuyo espectro extendido es de 10 Hz. La probabilidad de error por palabra es 1.4862×10^{-6} . El transmisor del multiplexor está operando linealmente con formato polar, con pulsos a la potencia pico; todo el ancho de banda del canal está siendo usado. El multiplexor se modifica para manejar el doble de mensajes similares y la modulación se cambia a PAM de 16 niveles (un nivel por cada palabra de 4 bits), cuya potencia pico es elevada 20 veces a la del sistema original.

a) ¿Cuál es la duración de trama de las 1000 señales multiplexadas?

b) ¿Cuál es la duración de un símbolo M-ARY?

c) ¿Cuál es la P_w del sistema?

d) ¿Por qué es necesario convertir a un sistema M-ARY PAM?

III.16. Una señal limitada en banda a 17.5 kHz debe ser reconstruida exactamente a partir de sus muestras. Si se muestrea en forma ideal y se recupera por medio de un filtro pasa bajas ideal:

a) ¿Cuál es el ancho de banda mínimo permitido para el filtro?

b) ¿Cuál es la frecuencia mínima a la que se debe muestrear la señal?

III.17. Una señal limitada en banda tiene una frecuencia máxima de 5 KHz y está siendo muestreada a una razón muy alta de muestreo de 100 KHz, usando pulsos ideales. En el receptor, la señal se filtra por un filtro pasa banda cuyo ancho de banda es de 15 KHz, centrado a una frecuencia de 100 KHz. ¿Se recupera la señal original sin distorsión utilizando este filtro? ¿Cómo es la forma de la señal $f_o(t)$ a la salida del filtro?

III.18. Se muestran 32 señales, cada una limitada a un ancho de banda de 3300 Hz y multiplexadas en tiempo a razón de 8 kHz. Calcule el ancho de banda mínimo necesario para transmitir esta señal multicanalizada en un sistema PAM con un tiempo de guarda igual a cero.

III.19. Se muestran cinco señales a la misma frecuencia y multiplexan en tiempo. La señal multiplexada es pasada por un filtro pasa bajas. Una señal tiene un ancho de banda de 5 KHz, dos de las señales tienen un ancho de banda de 3300 Hz y las otras dos señales restantes tienen un ancho de banda de 10 KHz.

a) ¿Cuál es la frecuencia mínima de muestreo?

b) Para esta frecuencia de muestreo, ¿cuál es el ancho de banda mínimo del filtro pasa bajas?

c) En el receptor las señales se demultiplexan y recuperan a través de filtros individuales, ¿cuál es el ancho de banda mínimo para cada uno de los filtros?

III.20. La señal $v(t) = \cos[5\pi t] + 0.5 \cos[10\pi t]$ instantáneamente se muestrea. El intervalo entre muestras es de T_s .

a) Calcule el valor máximo permitido para T_s .

b) Para reconstruir la señal $v(t)$ se pasa por un filtro pasa bajas ideal. Calcule el ancho de banda mínimo del filtro para reconstruir la señal sin distorsión.

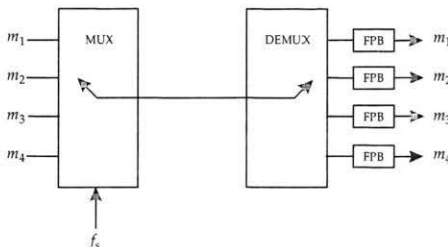
III.21. Una señal senoidal a 1 Hz, sen $2\pi t$ se muestrea periódicamente.

- a) Encuentre el intervalo máximo permisible para las muestras.
- b) Las muestras se toman a intervalos de $1/3$ seg. Efectúe la operación de muestreo gráficamente, y pruebe que ninguna otra función senoidal (o cualquier otra función temporal) de ancho de banda menor que 1.0 Hz puede ser representada por estas muestras.
- c) Las muestras se toman cada $2/3$ seg. Pruebe gráficamente que estas muestras pueden representar otra señal senoidal de frecuencia menor que 1.0 Hz.
- III.22. Una computadora transmite símbolos binarios a una razón de 64 Kbits/seg.
- a) ¿Cuál es el ancho de banda mínimo del canal de transmisión para transmitir esta señal?
- b) ¿Qué cambios deben hacerse en el sistema de transmisión para poder transmitir 64 Kbits/seg, sobre un canal de transmisión con un ancho de banda igual al encontrado en el inciso a) si el receptor usa un filtro cosenoidal?
- III.23. Una señal $m_1(t)$ es de banda limitada a 3 KHz y otras tres señales $m_2(t)$, $m_3(t)$ y $m_4(t)$ son de banda limitada a 1 kHz cada una. Estas señales se van a muestrear a la frecuencia de Nyquist. Las cuatro señales son multiplexadas en tiempo.
- a) Determine la frecuencia de Nyquist.
- b) Calcule el ancho de banda mínimo de la señal multiplexada. Explique por qué éste es el BW mínimo.
- III.24. El sistema TDM que aparece en la siguiente figura se usa para multiplexar cuatro señales:

$$\begin{aligned} m_1(t) &= 1.0 \cos [\omega_0 t] \\ m_2(t) &= 2.0 \cos [2 \omega_0 t] \\ m_3(t) &= 0.5 \cos [\omega_0 t] \\ m_4(t) &= 4.0 \cos [4 \omega_0 t] \end{aligned}$$

con $\omega_0 = 6.28 \times 10^3$ rad/seg.

- a) Calcule el BW mínimo a la salida del MUX.
- b) Si todas las señales se muestrean a la misma razón, calcule la frecuencia de muestreo mínima (en Hz).
- c) Determine la frecuencia de corte, mínima (en Hz) de cada uno de los filtros de recepción.



- III.25. Un módem transmite pulsos binarios a través de un canal telefónico cuyo ancho de banda es de 4 KHz.
- a) ¿Cuál es el número máximo de bits/seg que se pueden transmitir en estas condiciones? Justifique su respuesta.
- b) Si se desea transmitir 16 Kbits/seg sobre el mismo canal de transmisión, ¿qué cambios deben hacerse al módem? Justifique su respuesta.

BIBLIOGRAFÍA

- [1] Shanmugan, *Digital and Analog Communications Systems*, John Wiley.
- [2] Feher, K., *Digital Communications-Microwave Applications*, Prentice Hall.
- [3] Shannon, C. E., *A Mathematical Theory of Communications*, Bell Syst, Tech, J. 1948-27.
- [4] Miller, J. M. y V. S. Ahamed, *Digital Transmission Systems and Networks*, Computer Science Press.
- [5] Peyton, Z. P., *Digital Communication Systems*, Prentice Hall.
- [6] Schwartz, Misha, *Information Transmission, Modulation and Noise*, McGraw-Hill.
- [7] Taub, H. y L. D. Schilling, *Principles of Communication Systems*, McGraw-Hill.

CAPÍTULO IV

TRANSMISIÓN DIGITAL EN BANDA BASE

IV.1. INTRODUCCIÓN

EL PROPÓSITO DE ESTE CAPÍTULO es describir las diversas funciones de la figura III.3 que están relacionadas con los sistemas de comunicación digital de banda base. Principalmente nos enfocaremos al convertidor analógico a digital, y con las funciones de los codificadores de canal y de fuente. Las funciones del subsistema de recepción las analizaremos muy brevemente, ya que básicamente son las funciones inversas del subsistema de transmisión.

Trataremos principalmente de un número de importantes formas de onda para transmisión sobre el canal; sin embargo, debido a que una de las grandes ventajas de los sistemas de comunicación digital es su habilidad para intercalar en tiempo, o multiplexar en tiempo, formas de ondas digitales de diversos mensajes, también se analizarán los elementos básicos de esta técnica.

Puesto que la conversión analógico a digital es la única función en el subsistema de transmisión que involucra formas de onda analógicas, además de que la teoría de muestreo es el proceso inicial de esta función (analizada en el capítulo anterior), comenzaremos con este tópico, para después continuar con los conceptos digitales.

IV.2. MODULACIÓN POR PULSOS CODIFICADOS (PCM)

En el capítulo anterior vimos que un mensaje analógico, en ciertas condiciones, puede muestrearse y transportar íntegramente la información original. Sin embargo, debe observarse que la señal muestreada no es una señal digital, ya que la amplitud de los pulsos puede tomar infinitos valores dentro de su rango de existencia.

También en el capítulo I dijimos que la utilización de señales digitales tiene muchos atractivos, por lo que es conveniente digitalizar la señal analógica. El proceso con el que una señal muestreada se digitaliza recibe el nombre de *cuantificación* y consiste en transformar los niveles de amplitud continuos de la señal de entrada en un conjunto de niveles discretos previamente establecidos. En otras palabras, la cuantificación aproxima o redondea las amplitudes de las muestras en un cierto número de niveles preestablecidos.

En el caso de señales muestreadas, la transmisión está compuesta por información analógica enviada en tiempos discretos; es decir, en una señal muestreada la variación de la amplitud de los pulsos puede variar en forma continua sobre todos los valores. Un refinamiento adicional es cuantificar la señal analógica muestreada en cierto número de niveles discretos, o sea, ahora no sólo se cuantifica el tiempo sino también las amplitudes de la señal; si adicionalmente a cada una de estas amplitudes discretas se les asigna un código para designar cada nivel, se obtiene un tipo de modulación conocido como *modulación por pulsos codificados* (PCM). De ahí que este proceso de modulación

involucre varias operaciones: muestreo, cuantificación y codificación, que serán comentadas en detalle más adelante.

A diferencia del muestreo, el proceso de cuantificación implica una pérdida irremediable de información ya que resultará imposible reconstruir la señal analógica original a partir de la señal cuantificada. La discrepancia entre la señal de entrada y su versión cuantificada comúnmente se conoce como *ruido de cuantificación*. Esta denominación no es en realidad muy adecuada pues la mencionada degradación está fuertemente correlacionada con la señal, sin presentar en consecuencia las características de aleatoriedad que normalmente caracterizan al ruido. Más aún, este tipo de ruido cesa al anularse la señal de entrada, por lo que se considera como una distorsión. Algunos autores también suelen denominarlo *error de cuantificación*, y si bien nosotros preferimos la denominación de *distorsión de cuantificación* debe quedar claro que las tres denominaciones se refieren a lo mismo.

Cabe preguntar si las ventajas de la digitalización son tantas que se justifique introducir desde el inicio mismo de la transmisión una distorsión irreversible. La respuesta es indudablemente afirmativa, sobre todo porque esa distorsión es controlable mientras que las distorsiones que sufren las señales analógicas no sólo son irrecuperables sino en la mayoría de los casos imposibles de acotar. Para decirlo más coloquialmente: “es preferible pagar distorsión de cuantificación al contado que ruido en cuotas”.

Como podemos apreciar, el tema central del proceso de cuantificación es la evaluación y, en consecuencia, la minimización de la distorsión de cuantificación. Esa distorsión, expresada ya sea en potencia o como valor *rms*, en realidad no es una medida ideal de la degradación tal como la juzgaría subjetivamente un observador que escucha una voz cuantificada o que observa una imagen cuantificada. Con todo, es de mucha utilidad para el diseño y la especificación de los sistemas, habida cuenta que el valor recomendado debe determinarse subjetivamente y no por analogía con el ruido aleatorio ordinario.

El proceso de cuantificación puede o no ser uniforme, y si bien el segundo es de aplicación práctica en los sistemas telefónicos, comenzaremos describiendo el caso uniforme por razones de simplicidad.

IV.2.1. El proceso de cuantificación

El proceso de cuantificación, también llamado muestreo en amplitud, consiste en dividir el rango de amplitudes de la señal original en intervalos uniformes; y todas las muestras cuya amplitud caiga dentro de un intervalo específico de cuantificación se les asignará como se muestra en la figura IV.1 el mismo valor de amplitud a la salida.

Los valores discretos a los que se redondean las muestras se denominan niveles de cuantificación, y se llama intervalo de cuantificación al intervalo de captura correspondiente.

En la figura IV.1 vemos una onda senoidal procesada por un cuantificador uniforme; es decir, los intervalos de cuantificación son todos del mismo valor. Evidentemente la distorsión por cuantificación está limitada a

$$|e_q(t)|_{\max} \leq d/2 \quad (\text{IV.1})$$

donde $e_q(t)$ es la distorsión de cuantificación, que analizaremos más adelante con mayor detalle, así como el cálculo de los niveles óptimos de cuantificación.

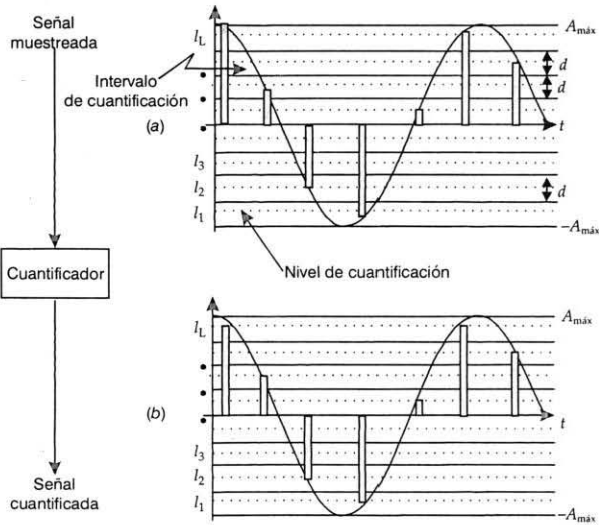


Figura IV.1. Proceso de cuantificación. a) Señal muestreada, b) señal cuantificada

IV.2.2. El proceso de codificación

La operación final para obtener una señal modulada en PCM es la codificación, que consiste en asignar a cada nivel de cuantificación una representación numérica, la mayoría de las veces binaria; este proceso se ilustra en la figura IV.2, donde se supone que se están utilizando ocho niveles de cuantificación; además, se muestran dos distintos códigos de línea que se pueden utilizar para la transmisión de la señal PCM. Las técnicas de codificación de canal las estudiaremos en la sección IV.6.

En el proceso de la figura IV.2a se muestra una codificación binaria que se utiliza muchas veces en los convertidores analógicos a digitales, denominada código binario con signo; en b) y c) aparecen dos posibles codificaciones de canal conocidas como unipolar y codificación alternada de marca (AMI), en ambas figuras se aprecia con facilidad que lo que realmente se transmite es la palabra de código que representa la amplitud de la muestra; de ahí que se puede interpretar el porqué a este esquema de modulación se le conoce como modulación por pulsos codificados.

IV.2.3. Demodulación de señales PCM

Las operaciones que se efectúan en el subsistema de recepción para recuperar la señal analógica original, a partir de la señal PCM, son: regeneración de pulsos, decodificación y reconstrucción. Como mencionamos anteriormente, el proceso de cuantificación al ser un proceso de redondeo introduce un error irrecuperable, por lo que no es posible realizar un proceso inverso a éste. La regeneración consiste en recrear los pulsos transmitidos, distorsionados por el canal de transmisión tal como se explicó en el capítulo III; aunada a esta distorsión hay que tomar en cuenta la perturbación que produce el ruido en el canal. Entonces, este proceso consiste en muestrear los pulsos recibidos en los instantes adecuados para evitar la interferencia entre símbolos, y a partir

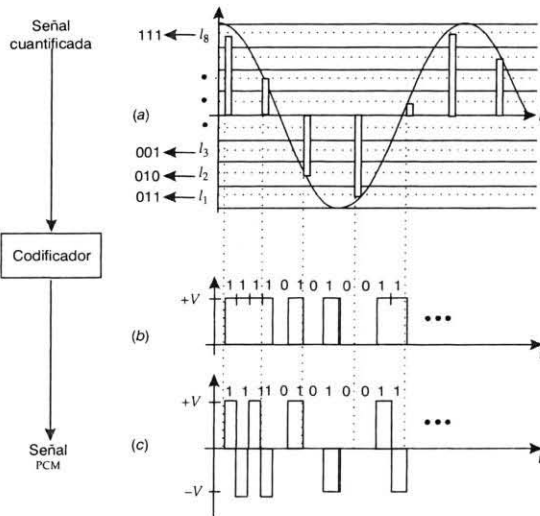


Figura IV.2. Proceso de codificación. a) Señal cuantificada, b) señal PCM transmitida en código unipolar y c) señal transmitida en código AMI

de un nivel de decisión (seleccionado para minimizar la probabilidad de error) que depende del código de línea utilizado, determinar la ausencia o presencia del pulso. En la sección IV.7 calculamos los niveles de decisión para varios códigos de línea; en la figura IV.3 se muestra el proceso de regeneración de pulsos.

Este proceso de regeneración de pulsos permite recrear los pulsos transmitidos a una forma idéntica a como los generó el transmisor, sin amplificar el ruido como en los repetidores utilizados en los sistemas de comunicación analógicos; esta característica permite que los sistemas digitales

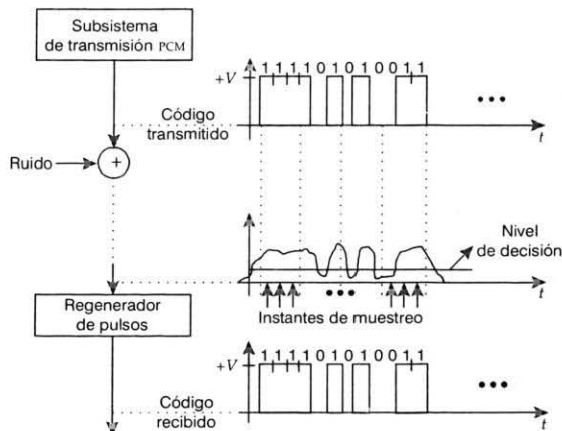


Figura IV.3. Proceso de regeneración de pulsos

puedan utilizarse para transmitir sobre grandes distancias utilizando a intervalos adecuados repetidores regenerativos, que reproducen fielmente la misma señal generada en el transmisor. Dicho en otras palabras: el parámetro de calidad en los sistemas de comunicación digital no es la relación señal a ruido sino más bien la probabilidad de cometer un error en el proceso de detección del pulso; el cálculo de la probabilidad de error se verá más adelante.

El proceso de decodificación es una operación inversa al proceso de codificación y básicamente consiste en ir reproduciendo las muestras adecuadas con base en el código recibido; el proceso de reconstrucción, por su parte, reconstruye la señal original a partir de estas muestras por medio de un proceso de filtraje, tal como se explicó en el capítulo III.

Ahora calcularemos el ancho de banda necesario de N señales PCM multiplexadas en tiempo, partiremos del ancho de banda de N señales multiplexadas en tiempo con tiempo de guarda igual a cero, que está dado por $B_{TDM} = N/T_s$; en este caso cada ranura de tiempo la ocupa totalmente una muestra; si tomamos en cuenta que en PCM cada muestra se codifica en general con n pulsos, entonces el ancho de banda se incrementa por un factor igual a n , es decir

$$B_{PCM} = \frac{N \cdot n}{T_s} = N \cdot n \cdot f_s \text{ (Hz)} \quad (IV.2)$$

De manera similar, es claro que el número de pulsos transmitidos por N señales PCM multiplexadas en tiempo es:

$$r = N \cdot n \cdot f_s \text{ (pulsos/seg)} \quad (IV.3)$$

Cabe aclarar que la velocidad es numéricamente igual al ancho de banda, siempre y cuando tenga unidades de pulsos/seg (bauds). Recordemos que en el capítulo II dijimos que un pulso (símbolo) puede transportar más de un bit de información; en el único caso en el que la velocidad en pulsos/seg (bauds) es igual a la velocidad en bit/seg, es cuando cada pulso transporta un bit de información, lo que corresponde al caso binario con símbolos equiprobables.

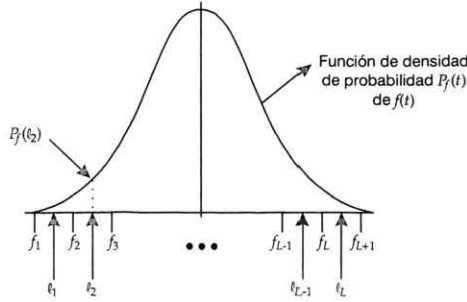
IV.3. DISTORSIÓN Y NIVELES ÓPTIMOS DE CUANTIFICACIÓN

Supongamos que la señal analógica $f(t)$ está modelada como una señal aleatoria. Definamos la función de densidad de probabilidades (apéndice B) de $f(t)$ a cualquier tiempo como $p_f(f)$, tal como se muestra en la figura IV.4. Como ya dijimos, el proceso de cuantificación subdivide el intervalo de valores de f en intervalos discretos. Si una muestra particular de $f(t)$ cae en cualquier intervalo, se le asigna un valor discreto correspondiente al intervalo. En la figura IV.4 a las fronteras de los intervalos se les identificó como $f_1, f_2 \dots f_{L+1}$, donde se supone que hay L intervalos de cuantificación; asimismo, a los niveles de cuantificación se les ha denotado como $l_1, l_2 \dots l_L$.

Consideremos una muestra f que cae en el intervalo i , esto es $f_i \leq f \leq f_{i+1}$, entonces la salida del cuantificador será el nivel l_i que difiere del valor de la muestra por el error

$$\epsilon_i = f - l_i \quad i = 1, 2 \dots L \quad (IV.4)$$

con esto notamos que los mensajes originales nunca pueden recuperarse. Cuando el receptor reconstruye el mensaje a partir de sus niveles de cuantificación, éste contendrá errores relacionados con los diversos errores ϵ_i que se producen durante la cuantificación. Estos errores establecen un límite al rendimiento del sistema en su totalidad.

Figura IV.4. Función de densidad de probabilidad de $f(t)$ y niveles de cuantificación

Estamos interesados en encontrar un *error de cuantificación cuadrático medio* total, denotado por ϵ_q^2 , que aparece debido al proceso de cuantificación. Supongamos que $p_f(f)$ representa la función de densidad de probabilidades de $f(t)$ y consideremos un cuantificador definido por los intervalos y niveles mostrados en la figura IV.4. El valor medio y la potencia de la señal analógica pueden escribirse:

$$\overline{f(t)} = \bar{f} = \int_{-\infty}^{\infty} f \cdot p_f(f) df = \sum_{i=1}^L \int_{f_i}^{f_{i+1}} f \cdot p_f(f) df \quad (\text{IV.5})$$

$$\overline{f^2(t)} = \bar{f}^2 = \int_{-\infty}^{\infty} f^2 \cdot p_f(f) df = \sum_{i=1}^L \int_{f_i}^{f_{i+1}} f^2 \cdot p_f(f) df \quad (\text{IV.6})$$

Para cualquier muestra, la salida del cuantificador puede tratarse como una variable aleatoria discreta, denotada por f_q , que puede tomar los niveles l_1, l_2, \dots, l_L . La probabilidad de que f_q pueda tomar un valor típico, por decir l_i , es justamente la probabilidad de que f caiga en el intervalo i , denotada por p_i . El valor medio y la potencia presente a la salida del cuantificador se obtienen promediando la variable aleatoria discreta sobre todos los niveles de cuantificación

$$\bar{f}_q = \sum_{i=1}^L p_i \cdot l_i = \sum_{i=1}^L l_i \cdot \int_{f_i}^{f_{i+1}} p_f(f) df \quad (\text{IV.7})$$

$$\bar{f}_q^2 = \sum_{i=1}^L l_i^2 \cdot p_i = \sum_{i=1}^L l_i^2 \cdot \int_{f_i}^{f_{i+1}} p_f(f) df \quad (\text{IV.8})$$

donde

$$p_i = \int_{f_i}^{f_{i+1}} p_f(f) df \quad (\text{IV.9})$$

Ahora, si se desea que el valor medio de la señal del cuantificador sea igual al valor medio de la señal original, es decir que en promedio las dos señales sean iguales, se debe cumplir que

$$\sum_{i=1}^L \int_{f_i}^{f_{i+1}} f \cdot p_f(f) df = \sum_{i=1}^L l_i \cdot \int_{f_i}^{f_{i+1}} p_f(f) df \quad (\text{IV.10})$$

para que las dos sumatorias sean iguales, la igualdad se debe cumplir término a término, lo que implica una condición más estricta, ya que ahora se está pidiendo que el valor medio de la señal cuantificada sea igual al valor medio de la señal original en cada uno de los intervalos de cuantificación; es decir, se debe cumplir que

$$\int_{f_i}^{f_{i+1}} f \cdot p_f(f) df = l_i \cdot \int_{f_i}^{f_{i+1}} p_f(f) df \quad i = 1, 2 \dots L \quad (\text{IV.11})$$

de donde, despejando l_i , los niveles de cuantificación óptimos están dados por

$$l_i = \frac{\int_{f_i}^{f_{i+1}} f \cdot p_f(f) df}{\int_{f_i}^{f_{i+1}} p_f(f) df} \quad i = 1, 2 \dots L \quad (\text{IV.12})$$

Es decir, para que se cumpla la condición $\bar{f} = \bar{f}_q$ los niveles de cuantificación deben ser iguales al valor medio condicional de cada intervalo.

Ejemplo IV.1

Calcule los niveles de cuantificación óptimos para el caso de que $p_f(f)$ sea constante; es decir, que sea igualmente probable que $f(t)$ tome cualquier valor dentro de su intervalo de amplitudes.

Suponiendo que $p_f(f) = k$, donde k es una constante, entonces el valor de las integrales de (IV.12) están dadas por

$$\begin{aligned} \int_{f_i}^{f_{i+1}} p_f(f) df &= k \cdot \int_{f_i}^{f_{i+1}} df = k \cdot (f_{i+1} - f_i) \\ \int_{f_i}^{f_{i+1}} f \cdot p_f(f) df &= k \cdot \int_{f_i}^{f_{i+1}} f df = k \cdot \left[\frac{(f_{i+1})^2 - (f_i)^2}{2} \right] \end{aligned}$$

luego, el valor de l_i es

$$l_i = \frac{[(f_{i+1})^2 - (f_i)^2]}{2[f_{i+1} - f_i]} = \frac{f_{i+1} + f_i}{2} \quad (\text{IV.13})$$

en la figura IV.5 se grafica la función de transferencia del cuantificador para $f_i = [-4, -3, -2, -1, 0, 1, 2, 3, 4]$; como se ve en la figura, los niveles de cuantificación óptimos que se obtuvieron utilizando (IV.13) caen exactamente a la mitad de cada intervalo de cuantificación, y para este ejemplo son $l_i = [-3.5, -2.5, -1.5, -0.5, 0.5, 1.5, 2.5, 3.5, 4.5]$.

Otro conjunto de niveles de cuantificación óptimos que se utilizan en la práctica son los que coinciden con las fronteras de los intervalos, es decir $l_i = f_i$; el cálculo de estos niveles y la función de transferencia del cuantificador los dejamos como ejercicio para el lector (véase problema 1).

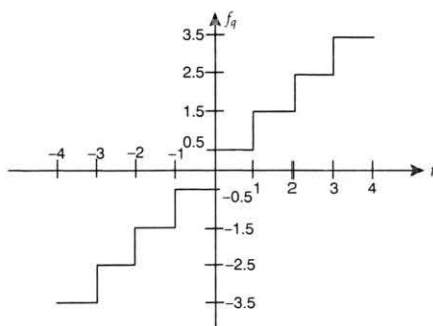


Figura IV.5. Función de transferencia del cuantificador

Retomando nuestro objetivo original que es calcular el error de cuantificación cuadrático medio, a partir de la (IV.4) el error cuadrático medio cuando la muestra cae en el intervalo i es

$$\overline{\epsilon_i^2} = \int_{f_i}^{f_{i+1}} (f - l_i)^2 \cdot p_f(f|i) df \quad (\text{IV.14})$$

donde $p_f(f|i)$ es la densidad de probabilidad de f cuando f cae en el intervalo i , la cual está dada por:

$$p_f(f|i) = \frac{p_f(f)}{p_i} \quad (\text{IV.15})$$

Promediando estos errores cuadráticos medios sobre todos los intervalos obtenemos el error cuadrático medio total

$$\overline{\epsilon_q^2} = \sum_{i=1}^L \overline{\epsilon_i^2} \cdot p_i = \sum_{i=1}^L \int_{f_i}^{f_{i+1}} (f - l_i)^2 p_f(f) df \quad (\text{IV.16})$$

expandiendo el lado derecho de (IV.16) obtenemos

$$\begin{aligned} \sum_{i=1}^L \left[\int_{f_i}^{f_{i+1}} f^2 p_f(f) df + \int_{f_i}^{f_{i+1}} l_i^2 p_f(f) df - 2 \cdot l_i \int_{f_i}^{f_{i+1}} f p_f(f) df \right] \\ = \overline{f^2} + \overline{f_q^2} - 2 \sum_{i=1}^L l_i \cdot l_i \int_{f_i}^{f_{i+1}} p_f(f) df = \overline{f^2} - \overline{f_q^2} \end{aligned}$$

es decir

$$\overline{\epsilon_q^2} = \overline{f^2} - \overline{f_q^2} \quad (\text{IV.17})$$

En otras palabras, el error cuadrático medio total de la distorsión por cuantificación es igual a la diferencia entre la potencia de la señal analógica de entrada y la potencia promedio de la señal de salida del cuantificador.

A partir de esto es posible definir una relación señal a ruido de cuantificación para la señal recuperada en el receptor. Esta relación está dada por

$$\frac{S_q}{N_q} = \frac{\overline{f_q^2}}{\overline{\epsilon_q^2}} = \frac{\overline{f^2} - \overline{\epsilon_q^2}}{\overline{\epsilon_q^2}} = \frac{\overline{f^2}}{\overline{\epsilon_q^2}} - 1 \quad (\text{IV.18})$$

donde se ha utilizado el resultado de la ecuación (IV.17) y $N_q = \overline{\epsilon_q^2}$ se trata como una potencia de ruido en la salida. Típicamente, $\frac{\overline{f^2}}{\overline{\epsilon_q^2}} \gg 1$ en sistemas prácticos, de tal manera que (IV.18) puede escribirse como

$$\left(\frac{S_q}{N_q} \right) \approx \left(\frac{S_o}{N_q} \right) = \frac{\overline{f^2}}{\overline{\epsilon_q^2}} \quad (\text{IV.19})$$

donde

$$S_o = \overline{f^2} = \overline{f^2(t)} \quad (\text{IV.20})$$

es la potencia de la señal original.

La ecuación (IV.19) nos indica que la relación señal a ruido de cuantificación es una simple relación de potencias entre la señal de entrada y la señal cuantificada; sin embargo, todavía falta obtener un valor numérico para la potencia de la distorsión por cuantificación. Para obtener este resultado partiremos de la ecuación (IV.16), que repetiremos aquí para mayor claridad

$$\overline{\epsilon_q^2} = \sum_{i=1}^L \overline{\epsilon_i^2} \cdot p_i = \sum_{i=1}^L \int_{f_i}^{f_{i+1}} (f - l_i)^2 p_f(f) df$$

Suponiendo que sea igualmente probable que la amplitud de la señal muestreada esté en cualquier punto dentro de un intervalo, entonces la función de densidad de probabilidad $p_f(f)$ debe ser constante, es decir $p_f(f) = k$, donde k es una constante, en estas condiciones la ecuación (IV.16) se puede escribir como

$$\overline{\varepsilon_q^2} = \sum_{i=1}^L k \cdot \int_{f_i}^{f_{i+1}} (f - l_i)^2 df \quad (\text{IV.21})$$

tomando la altura del intervalo de cuantificación igual a d , figura IV.1, podemos cambiar los límites de la integral a

$$\overline{\varepsilon_q^2} = \sum_{i=1}^L k \cdot \int_{f_i - d/2}^{f_{i+1} + d/2} (f - l_i)^2 df \quad (\text{IV.22})$$

haciendo el cambio de variable $x = f - l_i$ se obtiene

$$\overline{\varepsilon_q^2} = \sum_{i=1}^L k \cdot \int_{-d/2}^{d/2} x^2 dx = \sum_{i=1}^L k \left. \frac{x^3}{3} \right|_{-d/2}^{d/2} = \sum_{i=1}^L k \cdot \frac{d^3}{12} \quad (\text{IV.23})$$

Como la probabilidad de que la amplitud de la señal muestreada esté entre $f_i - d/2$ y $f_i + d/2$ es $k \cdot d$, entonces

$$\sum_{i=1}^L k \cdot d = 1 \quad (\text{IV.24})$$

luego, el valor cuadrático medio de la distorsión por cuantificación es:

$$\overline{\varepsilon_q^2} = \frac{d^2}{12} \quad (\text{IV.25})$$

Aunque la ecuación anterior ya nos permite obtener un valor numérico para la potencia del ruido de cuantificación, para calcular la relación señal a ruido de cuantificación a partir de la ecuación (IV.19) todavía necesitamos conocer la potencia de la señal de entrada; para propósitos de comparación se calculará esta relación para una señal senoidal en el siguiente ejemplo.

Ejemplo IV.2

Obtendremos la relación señal a ruido de cuantificación para a) una señal senoidal de amplitud pico igual a la máxima escala del cuantificador, y b) para el caso de que la amplitud pico de la senoidal sea menor a la máxima escala del cuantificador.

Para el caso a) podemos escribir la amplitud pico a pico de la señal senoidal como $V_{pp} = L \cdot d$, donde L es el número de niveles de cuantificación y d es el tamaño del intervalo, recordando que la

potencia efectiva de una onda senoidal está dada por $S_o = V_{pp}^2 / 8$, por lo que la relación señal a ruido de cuantificación para este caso es

$$\frac{S_o}{N_q} = \frac{L^2 d^2 / 8}{d^2 / 12} = \frac{3 L^2}{2} \quad (\text{IV.26})$$

Tomando en cuenta que L está relacionado con la longitud de la palabra de código, entonces para codificar un total de L niveles se necesitarán n dígitos, donde n y L están relacionados por $L = b^n$, donde b es la base del código utilizado. Sustituyendo la relación anterior obtenemos

$$\frac{S_o}{N_q} = 1.5 \cdot b^{2n} \quad (\text{IV.27})$$

expresando esta relación en decibelios, la ecuación (IV.27) se puede escribir como

$$\frac{S_o}{N_q} \text{ (dB)} = 10 [\log (1.5 \cdot b^{2n})] = 1.76 + 20n \log b \quad (\text{IV.28})$$

la expresión anterior es válida para cualquier base; si se utiliza un código binario $b = 2$ entonces se simplifica a

$$\frac{S_o}{N_q} \text{ (dB)} = 1.76 + 6.02 n \quad (\text{IV.29})$$

es decir, la relación se mejora en 6 dB por cada bit que se aumenta a la longitud de la palabra de código. Sin embargo, recordando el ancho de banda de una señal PCM ($B_{\text{PCM}} = N n f_s$), esta mejora se obtiene a expensas de un incremento en el ancho de banda.

Reemplazando la relación de ancho de banda para una señal PCM en la relación $L = b^n$, para el caso binario tenemos que

$$L = 2^{B/2f} \quad (\text{IV.30})$$

donde B es el ancho de banda de una sola señal PCM, suponiendo que se está muestreando a la frecuencia de Nyquist, donde $f = f_{\text{máx}}$, ya introducida en la ecuación (IV.26) nos da

$$\frac{S_o}{N_q} = 1.5 \cdot 2^{B/f} \quad (\text{IV.31})$$

la que nos muestra que la relación señal a ruido de cuantificación se incrementa con B/f , es decir, la expansión de ancho de banda. Relaciones similares se encuentran en otros sistemas de modulación como PDM y FM. No obstante, debido a la relación exponencial de la ecuación (IV.31) la modulación PCM es superior a estos sistemas en lo referente al intercambio de ancho de banda por relación señal a ruido. Debe señalarse, sin embargo, que habitualmente se trata de utilizar conservadoramente el ancho de banda, por lo que los valores de B/f se mantienen usualmente bajos. Así, para un sistema de 8 bits (256 niveles) tendremos una relación señal a ruido de cuantificación de 49.92 dB y un consecuente aumento relativo de 16 veces en el ancho de banda.

Para el caso *b)* la potencia de la señal senoidal puede escribirse como $S_o = A^2/2$, donde A es el voltaje pico de la senoidal, por lo que la relación señal a ruido de cuantificación expresada en decibeles es

$$\frac{S_o}{N_q} (\text{dB}) = 10 \cdot \log \left(\frac{A^2/2}{d^2/12} \right) = 7.18 + 20 \log (A/d) \quad (\text{IV.32})$$

tomando en cuenta que el tamaño del intervalo de cuantificación lo podemos escribir como

$$d = \frac{2 A_{\text{máx}}}{L} = \frac{2 A_{\text{máx}}}{2^n} \quad (\text{IV.33})$$

donde suponemos que $A_{\text{máx}}$ es la máxima escala del cuantificador y $L = 2^n$ (caso binario); sustituyendo la ecuación (IV.33) en la (IV.32) obtenemos

$$\frac{S_o}{N_q} (\text{dB}) = 7.78 + 20 \log \left(\frac{2^n A}{2 A_{\text{máx}}} \right) = 1.76 + 6.02 n + 20 \log \left(\frac{A}{A_{\text{máx}}} \right) \quad (\text{IV.34})$$

que coincide con la ecuación (IV.29) cuando $A = A_{\text{máx}}$. En resumen, la (IV.34) nos indica que si la amplitud pico de la señal no llega al nivel más alto del cuantificador, habrá un deterioro de la relación S_o/N_q puesto que disminuye la potencia de la señal. En la figura IV.6 se representa la ecuación (IV.34) para varios valores de n .

Ejemplo IV.3

Una señal se cuantifica y transmite codificando las muestras cuantificadas. Si cada muestra debe tener un error de cuantificación que no supere $\pm 1\%$ del valor pico a pico del cuantificador, calcule el número de dígitos binarios que debe contener cada palabra de código. Calcule el ancho de banda de la señal que resulte suponiendo que el espectro de la señal analógica no supera los 4 kHz.

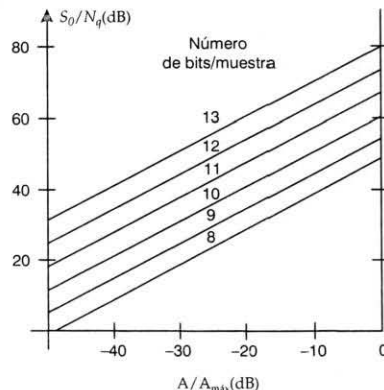


Figura IV.6. Relación señal-distorsión para un cuantificador uniforme

Puesto que el error máximo es $\pm 1\%$, cada intervalo de cuantificación debe ocupar 2% del valor pico a pico del cuantificador, lo que significa 50 niveles de cuantificación.

Tomando $n = 6$ dígitos binarios por palabra de código, tenemos:

$$L = 2^n = 64 \text{ niveles de cuantificación}$$

Según la ecuación (IV.2) el ancho de banda de transmisión será:

$$B_{\text{PCM}} = N n f_s = (1) (6) (2 \cdot 4 \text{ kHz}) = 48 \text{ kHz}.$$

Donde se supuso que se muestrea a la tasa de Nyquist.

Ejemplo IV.4

Continúe el problema anterior graficando el ancho de banda de transmisión B en función de la exactitud requerida.

Por exactitud ϵ queremos significar el máximo error relativo de cuantificación, para lo cual llamamos V al valor pico del cuantificador.

$$\epsilon = \frac{d/2}{2V} = \frac{d}{4V} = \frac{1}{2L}$$

puesto que $L = 2^{V/d}$, y como en el caso binario

$$L = 2^n = \frac{1}{2\epsilon}$$

o sea:

$$n = \log_2 (1/2\epsilon)$$

por lo que el ancho de banda de la señal es

$$B_{\text{PCM}} = N n f_s = 2 f_{\text{máx}} \log_2 [1/2\epsilon]$$

donde se supuso una sola señal PCM y que se muestrea a la tasa de Nyquist.

Graficando

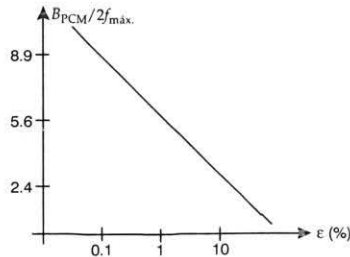


Figura IV.7. Variación del ancho de banda con la exactitud

Ejemplo IV.5

En un convertidor análogo digital (A/D) se desea digitalizar la señal $x(t) = 5 \cos [200 \pi t] + 7 \cos [600 \pi t]$. Determine:

- El rango dinámico del convertidor A/D.
- La relación S_o/N_q en función de la longitud de la palabra de código.
- De $x(t)$ se deduce que:

$$[x(t)]_{\max} = 12 \text{ y } [x(t)]_{\min} = -12$$

Por lo que el rango dinámico es $2A = 12 - (-12) = 24$

- El valor del error cuadrático medio de la distorsión por cuantificación suponiendo código binario es

$$N_q = \frac{d^2}{12} = \frac{(24/2^n)^2}{12} = 48 \cdot 2^{-2n}$$

la potencia de la señal $x(t)$ está dada por:

$$S_o = \frac{5^2}{2} + \frac{7^2}{2} = 37$$

En consecuencia

$$\frac{S_o}{N_q} = \frac{37}{48} \cdot 2^{2n}$$

$$S_o/N_q \text{ (dB)} = -1.13 + 6.02 n$$

Obsérvese que hay un desempeño de 2.89 dB ($= 1.76 + 1.13$) inferior a una señal senoidal debido a la diferencia de los factores de cresta.

Ejemplo IV.6

Estime el número mínimo de niveles (o bits) necesarios para obtener una calidad aceptable en un sistema telefónico. Establezcamos que la calidad mínima requerida la fijamos en 26 dB.

Hagamos el cálculo para el caso de señales senoidales que si bien no es el caso real, por su simplicidad nos ayudará a fijar conceptos y órdenes de magnitud. Supongamos primero que la señal senoidal de entrada ocupa todo el rango del cuantificador.

La relación S_o/N_q es

$$\frac{S_o}{N_q} \text{ (dB)} = 1.76 + 6.02 n$$

que para ser mayor que 26 dB requiere aproximadamente 4 bits (16 niveles). Sin embargo, para proveer una calidad aceptable en las señales pequeñas, el sistema telefónico debe proveer la calidad señalada dentro de un rango dinámico de 30 dB. Es decir, que señales 31 veces menores ($30 \approx 20 \log 31$) deben también cuantificarse en 16 niveles, con lo que el rango total del cuantificador requerirá $31 \times 16 = 496$ niveles, lo que corresponde a 9 bits de codificación.

Si tenemos en cuenta la ocurrencia de otras degradaciones y que en general hay varios procesos de conversión análogo-digital, nuestra conclusión es que necesitamos usar 12 o 13 bits cuando pretendamos seguir utilizando cuantificación uniforme.

Podremos profundizar más el análisis suponiendo distintas funciones de densidad de probabilidad y calculando para cada una de ellas la cantidad de bits necesarios. Sin embargo, los resultados hasta ahora obtenidos muestran que el número de bits es demasiado alto, y como vimos en la ecuación (IV.34) la situación empeora para señales más realistas, por lo que lo más conveniente es reservar el estudio detallado para el caso de cuantificación no uniforme.

IV.4. CUANTIFICACIÓN NO UNIFORME

El mayor inconveniente de la cuantificación uniforme radica en la constancia del error de cuantificación, lo que hace que la relación S_o/N_q disminuya al disminuir la señal de entrada tal como se evidencia en la figura IV.6. Como hemos visto, el problema particularmente serio es para las señales de alto factor de cresta que son precisamente las señales habituales en el caso telefónico. Sumado a ello debemos tener en cuenta que los distintos abonados presentan niveles vocales muy variados. Así, abonados de alto nivel sonoro pueden estar enlazados con líneas sumamente cortas y otros de bajo nivel sonoro pueden estar en el extremo de un circuito transcontinental. Si bien dentro de la red se toman las medidas tendientes a minimizar esta variación, se deben esperar dispersiones entre los volúmenes medios de hasta 30 dB. Partiendo de la premisa de que cada conexión debe asegurar una calidad aceptable, es necesario asignar intervalos de cuantificación que la provean, incluso para los volúmenes más bajos.

Esto demandaría un elevado número de bits por muestra, por lo que ningún sistema telefónico puede utilizar racionalmente cuantificación uniforme. Lo que deseamos es un cuantificador que provea una relación S_o/N_q constante dentro del rango dinámico de la señal de entrada.

Para ello, la distorsión de cuantificación debe disminuir para las señales de entrada más bajas y (en principio) debe hacerse proveyendo estructuras de cuantificación más finas cuanto menor sea el nivel de señal, o sea, comprimir los niveles de cuantificación más próximos al cero y expandir los extremos, por lo que el error de cuantificación será menor para los niveles más bajos y mayor para los más altos, lo que distribuye su efecto perjudicial. Este procedimiento de cuantificación en el que los intervalos de cuantificación no están igualmente espaciados, recibe el nombre de cuantificación no uniforme.

Es interesante notar que al cuantificar no uniformemente obtenemos una ventaja adicional a la ya mencionada relación señal-distorsión de cuantificación aceptable en todo el intervalo. En efecto, la distribución de las amplitudes de la señal de entrada tiene una función de densidad de probabilidad (f_{dp}) muy distante de ser uniforme, teniendo mayor probabilidad de ocurrencia para los bajos niveles que para los altos. Por ello, cuando se usa cuantificación uniforme las probabilidades asociadas con cada nivel son muy diferentes, por lo que las palabras de código que representan a cada nivel tampoco son equiprobables. Puesto que la señal codificada presenta la máxima entropía sólo cuando todos los niveles son equiprobables, resulta beneficioso asignar intervalos de cuantificación más grandes a los intervalos de la señal donde la f_{dp} es menor, y viceversa. Con ello se tiende a uniformar la probabilidad de captura de cada intervalo. En general, la maximización de la entropía no coincide con la relación S_o/N_q óptima buscada, pero las leyes que se utilizan para desuniformizar son cualitativamente coincidentes.

En resumen: en transmisión de voz se requiere que la relación señal a ruido de cuantificación sea mayor a 30 dB, y al mismo tiempo que para este valor de señal a ruido, la potencia de la señal

de entrada varíe 40 dB (es decir que la señal tenga un rango dinámico de 40 dB). A partir de la ecuación (IV.34) o de la figura IV.6 se deduce que para cumplir con estas condiciones el número de bits utilizados para la codificación de las muestras utilizando cuantificación uniforme debe ser por lo menos de 12; esto aumenta demasiado el ancho de banda de transmisión, por lo que resulta necesario tener un proceso de compansión.

Compansión es el proceso de comprimir y después expandir. Con los sistemas compandidos, las señales analógicas de amplitud más alta se comprimen (amplificadas menos que las señales de amplitud menor), antes de su transmisión, después expandidas (amplificadas más que las señales de amplitud más pequeñas) en el receptor.

La figura IV.8 ilustra el proceso de compansión. Una señal de entrada con un rango dinámico de 120 dB se comprime a 60 dB para transmisión, después se expande a 120 dB en el receptor. Con PCM, la compansión se puede lograr por medio de técnicas analógicas o digitales. Los primeros sistemas PCM utilizaban compansión analógica, mientras que los sistemas más modernos utilizan compansión digital.

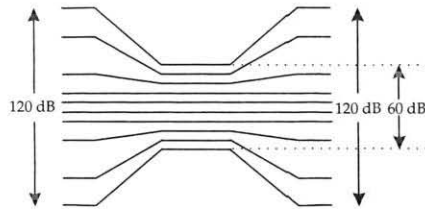


Figura IV.8. Procedimiento de compansión básico

IV.4.1. Compansión analógica

Históricamente, la compansión analógica se hizo usando diodos especialmente diseñados insertados en el camino de la señal analógica en el transmisor PCM, antes del circuito de muestreo y retención. La expansión analógica también se hizo con diodos colocados justo después del filtro pasa bajas de recepción. La figura IV.9 muestra el proceso básico de compansión analógico. En el transmisor, la señal analógica se comprime, se muestrea y después es convertida a un código PCM lineal. En el receptor, el código PCM se convierte a una señal PAM, la cual es filtrada, para después nuevamente ser expandido a sus características de amplitud de entrada originales.

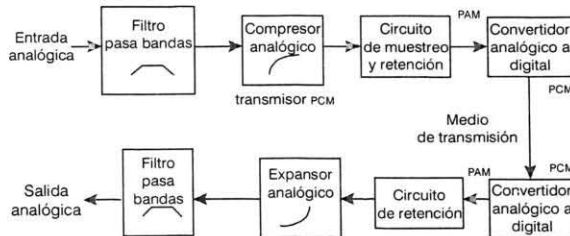


Figura IV.9. Sistema PCM con compansión analógica

Diferentes distribuciones de señal requieren diferentes características de compansión. Por ejemplo, las señales de voz requieren un rendimiento S_o/N_q relativamente constante sobre un amplio rango dinámico, lo que significa que la distorsión debe ser proporcional a la amplitud de la señal para cualquier nivel de señal de entrada; esto requiere una razón de compresión logarítmica. Un verdadero código de asignación logarítmica requiere un rango dinámico infinito y un número infinito de códigos PCM, lo que es imposible. Hay dos métodos de compansión analógica que actualmente se usan y, de manera cercana, se aproximan a una función logarítmica, que frecuentemente se llaman códigos log-PCM. Se trata de compansión de la ley μ y ley A .

IV.4.2. Leyes de compansión

En los Estados Unidos y Japón se usa la compansión de ley μ , su característica de compresión es

$$V_{\text{salida}} = \frac{V_{\text{máximo}} \cdot \ln(1 + \mu V_{\text{entrada}}/V_{\text{máximo}})}{\ln(1 + \mu)} \quad (\text{IV.35})$$

en donde:

$V_{\text{máximo}}$ = Máxima amplitud de entrada analógica descomprimida

V_{entrada} = Amplitud de la señal de entrada en un instante particular del tiempo

μ = Parámetro usado para definir la cantidad de compresión

V_{salida} = Amplitud de salida comprimida

La figura IV.10 muestra la compresión para varios valores de μ . Observe que entre más alta sea la μ hay más compresión. Además, observe que para que $\mu = 0$, la curva debe ser lineal (esto es, sin compresión).

El parámetro μ determina el rango de la potencia de la señal en la que la relación S_o/N_q es relativamente constante. Como ya dijimos, la transmisión de voz requiere un mínimo de rango dinámico de 40 dB y un código PCM de 7 bits; para un relación S_o/N_q relativamente constante y un rango dinámico de 40 dB, se requiere $\mu = 100$ o más. Los primeros sistemas de transmisión digital de Bell System usaban un código PCM de 7 bits con $\mu = 100$. Los sistemas de transmisión digitales más recientes utilizan códigos PCM de 8 bits y $\mu = 255$.

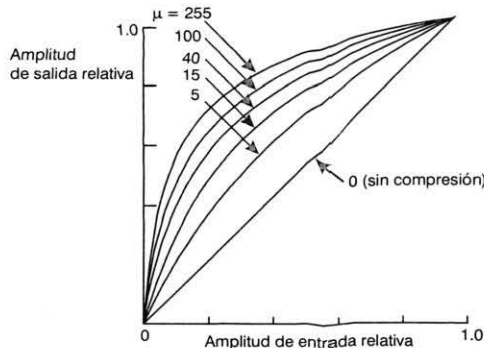


Figura IV.10. Características de compresión de la ley μ

Ejemplo IV.7

Para un compresor con $\mu = 255$, determine la ganancia para los siguientes valores de V_{entrada} : $V_{\text{máximo}}$, $0.75 V_{\text{máximo}}$, $0.5 V_{\text{máximo}}$ y $0.25 V_{\text{máximo}}$.

Solución: sustituyendo en la ecuación (IV.35), las siguientes ganancias se logran para varias magnitudes de entrada.

V_{entrada}	Ganancia
$V_{\text{máximo}}$	1.0
$0.75 V_{\text{máximo}}$	1.26
$0.5 V_{\text{máximo}}$	1.75
$0.25 V_{\text{máximo}}$	3.0

Puede observarse que conforme la amplitud de la señal de entrada se incrementa, la ganancia disminuye o se comprime.

En Europa, la ITT ha establecido la compansión de *ley A* para usarse y aproximarse al compansor logarítmico verdadero. Para un rango dinámico determinado, la compansión de *ley A* tiene una relación S_o/N_q ligeramente más plana que la *ley μ* . La compansión de *ley A*, sin embargo, es inferior a la *ley μ* en términos de calidad de señal pequeña (ruido de canal inactivo). La característica de compresión para compansión de *ley A* es

$$V_{\text{salida}} = V_{\text{máximo}} \frac{A (V_{\text{entrada}}/V_{\text{máximo}})}{1 + \ln A} \quad 0 \leq \frac{V_{\text{entrada}}}{V_{\text{máximo}}} \leq \frac{1}{A} \quad (\text{IV.36a})$$

$$V_{\text{salida}} = V_{\text{máximo}} \frac{1 + \ln (A V_{\text{entrada}}/V_{\text{máximo}})}{1 + \ln A} \quad \frac{1}{A} \leq \frac{V_{\text{entrada}}}{V_{\text{máximo}}} \leq 1 \quad (\text{IV.36b})$$

IV.4.3. Compansión digital

La compansión digital involucra la compresión por el lado de transmisión, después de que la muestra de entrada ha sido convertida a un código PCM lineal, y la expansión, en el lado de recepción, antes de la decodificación PCM. La figura IV.11 muestra el diagrama a bloques de un sistema PCM compansado de manera digital.

Con la compansión digital, la señal analógica primero se muestrea y se convierte a un código lineal, después éste se comprime de manera digital. Por el lado de recepción, el código PCM recibido comprimido, se expande y después se decodifica. Los sistemas PCM comprimidos de manera digital, más recientes, utilizan un código lineal de 12 bits y un código comprimido de 8 bits. Este proceso de compansión se asemeja a una curva de compresión analógica $\mu = 255$, aproximando la curva con un conjunto de ocho segmentos de línea recta (segmentos 0 a 7). La pendiente de cada segmento sucesivo es exactamente la mitad que el segmento previo. La figura IV.12 muestra la curva de compresión digital, de 12 bits a 8 bits, sólo para los valores positivos. La curva para los valores negativos es idéntica, excepto a la inversa. Aunque hay 16 segmentos (8 positivos y 8 negativos) este esquema frecuentemente se llama compresión a 13 segmentos. Esto se debe a que la curva para los segmentos +0, +1, -0 y -1 es una línea recta con una pendiente constante y frecuentemente se considera como un segmento.

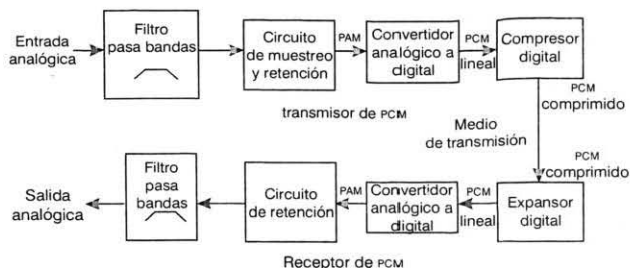


Figura IV.11. Sistema PCM compandido digitalmente

El algoritmo de compansión digital para un código comprimido, de 12 bits lineal a 8 bits, en realidad es muy sencillo. El código comprimido de 8 bits se compone de un bit de signo, un identificador de segmento de 3 bits y un código de magnitud de 4 bits que identifica al intervalo de cuantificación dentro del segmento especificado (véanse los cuadros IV.1, IV.2 y IV.3).

En el cuadro de codificación $\mu 255$ (cuadro IV.2), las posiciones del bit designadas con una x se truncan durante la compresión y son consecuentemente pérdidas. Los bits designados A, B, C y D se transmiten como están, lo mismo que el bit de signo (s). Observe que para los segmentos 0 y 1, los 12 bits originales se duplican exactamente a la salida del decodificador (cuadro IV.3), mientras que para el segmento 7 sólo los 6 bits más significativos se recuperan. Con 11 bits de magnitud hay 2 048 códigos posibles. Hay 16 códigos en el segmento 0 y en el segmento 1; en el 2 hay 32 códigos; el segmento 3 tiene 64. Cada segmento sucesivo, comenzando con el segmento 3, tiene el doble de la cantidad de códigos que el segmento anterior. En cada uno de los ocho segmentos sólo se pueden recuperar 16 códigos de 12 bits. En consecuencia, en los segmentos 0 y 1 no hay compresión (de los 16 códigos posibles, los 16 se pueden recuperar). En el segmento 2 hay una razón de compresión de 2:1 (32 códigos de transmisión posibles y 16 códigos recuperados posibles). En el segmento 3 hay una razón de compresión de 4:1 (64 códigos de transmisión posibles y 16 códigos recuperados posibles). La razón de compresión se duplica con cada segmento sucesivo. La razón de compresión en el segmento 7 es 1 024/16 o 64:1.

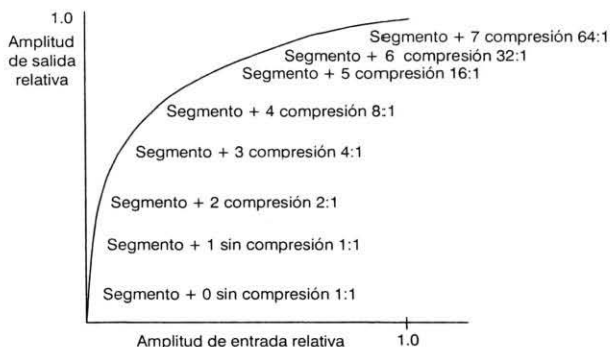


Figura IV.12. Características de compresión $\mu 255$ (solamente valores positivos)

CUADRO IV.1. Formato de código comprimido $\mu 255$ de 8 bits

Bit de signo 1 = + 0 = -	Identificador de segmento de tres bits 000 a 111	Intervalo de cuantificación de 4 bits A B C D 0000 a 1111
--------------------------------	---	---

CUADRO IV.2. Codificación $\mu 255$

Segmento	Código lineal de 12 bits	Código comprimido de 8 bits
0	s0000000ABCD	s000ABCD
1	s0000001ABCD	s001ABCD
2	s000001ABCDX	s010ABCD
3	s00001ABCDXX	s011ABCD
4	s0001ABCDXXX	s100ABCD
5	s001ABCDXXXX	s101ABCD
6	s01ABCDXXXXX	s110ABCD
7	s1ABCDXXXXXX	s111ABCD

CUADRO IV.3. Decodificación $\mu 255$

Código comprimido de 8 bits	Código recuperado de 12 bits	Segmento
s000ABCD	s0000000ABCD	0
s001ABCD	s0000001ABCD	1
s010ABCD	s000001ABCD1	2
s011ABCD	s00001ABCD10	3
s100ABCD	s0001ABCD100	4
s101ABCD	s001ABCD1000	5
s110ABCD	s01ABCD10000	6
s111ABCD	s1ABCD100000	7

El proceso de compresión es de la siguiente manera: la señal analógica se muestrea y se convierte a un código de magnitud de signo de 12 bits. El bit de signo se transfiere directamente al código de 8 bits. El segmento se determina contando el número de ceros principales en la porción de magnitud de 11 bits del código, comenzando con el MSB. Reste el número de ceros principales (no exceda de 7) de 7. El resultado es el número de segmento, el cual se convierte a un número binario de 3 bits y se sustituye en el código de 8 bits como el identificador del segmento. Los cuatro bits de magnitud (A, B, C y D) son el intervalo de cuantificación y se sustituyen en los 4 bits menos significativos del código comprimido de 8 bits.

Esencialmente, los segmentos 2 a 7 se subdividen en subsegmentos más pequeños. Cada segmento tiene 16 subsegmentos, que corresponden a las 16 condiciones posibles para los bits *A*, *B*, *C* y *D* (0000-1111). En el segmento 2 hay dos códigos por subsegmento. En el segmento 3 hay cuatro. El número de códigos por subsegmento se duplica con cada segmento subsecuente. Consecuentemente, en el segmento 7 cada subsegmento tiene 64 códigos. Observe que en cada subsegmento todos los códigos de 12 bits, una vez comprimidos y expandidos, se convierten en un solo código de 12 bits.

El más significativo de los bits truncados se reinserta en el decodificador como un 1. Los bits truncados restantes se reinsertan como ceros. Esto asegura que la máxima magnitud de error introducida por el proceso de compresión y expansión se minimiza. Esencialmente, el decodificador adivina que eran los bits truncados antes de codificar. La conjetura más lógica está a la mitad, entre el mínimo y el máximo código de magnitud. Por ejemplo, en el segmento 5, los 5 bits menos significativos se truncan durante la compresión. En el receptor, el decodificador debe determinar lo que esos bits eran. Las posibilidades son cualquier código entre 00000 y 11111. La conjetura lógica es 10000, aproximadamente la mitad de la máxima magnitud. En consecuencia, el máximo error de compresión es ligeramente más que la mitad de la magnitud de ese segmento.

Ejemplo IV.8

Para una resolución de 0.01 V y voltajes de la muestra analógicas de a) 0.05 V, b) 0.32 V y c) 10.23 V, determine el código lineal de 12 bits, el código comprimido de 8 bits y el código de 12 bits recuperado.

a) Para determinar el código lineal de 12 bits para 0.05 V, simplemente se divide el voltaje de la muestra por la resolución y el resultado se convierte a un número binario de magnitud de signo de 12 bits.

Código lineal de 12 bits:

$$\frac{0.05 \text{ V}}{0.01} = 5 = \begin{matrix} 10000000101 \\ \text{s—magnitud—} \end{matrix}$$

(número binario de 11 bits)

Código comprimido de 8 bits

1	0 0 0 0 0 0	0 1 0 1
s	(7 - 7 = 0 0 00)	A B C D
1	0 0 0	0 1 0 1
bit de signo (+)	Identificador unitario (segmento 0)	Intervalo de cuantificación

Código recuperado de 12 bits

1	0 0 0	0 1 0 1
s	(7 - 0 = 7 ceros principales)	A B C D
1	0 0 0 0 0 0 0	0 1 0 1
bit de signo (+)	Identificador del segmento determina el número de ceros	Intervalo de cuantificación

Como puede verse, el código de 12 bits recuperado es exactamente igual al código lineal de 12 bits original. Esto es cierto para todos los códigos en el segmento 0 y 1. Consecuentemente, no hay error de compresión en estos dos segmentos.

b) Para la muestra 0.32 V:

Código lineal de 12 bits:

$$\frac{0.32 \text{ V}}{0.01} = 32 = \begin{matrix} 100000100000 \\ \text{s—magnitud—} \end{matrix}$$

Código comprimido de 8 bits

		No transmitido	
1	. 0 0 0 0 0 1	0 0 0 0 0	
s	(7 - 5 = 2 o 010)	A B C D X	
1	0 1 0	0 0 0 0	↑
(+)	(segmento 2)		truncado

Código recuperado de 12 bits

1	0 1 0	0 0 0 0
s	(7 - 2 = 5 ceros principales)	A B C D X
1	0 0 0 0 0 1	0 0 0 0 1
	↑	↑
	insertado	insertado

Observe los dos 1 insertados en el código de 12 bits decodificado. El bit menos significativo se determina con el cuadro de decodificación. El 1 en la posición de bit 6 se eliminó durante la conversión de 12 bits a 8 bits. La transmisión de este bit es redundante, porque si no fuera un 1, la muestra no estaría en el segmento 3. Consecuentemente, en todos los segmentos, excepto 0, se inserta un 1 automáticamente después de los ceros reinsertados. Para este ejemplo, hay un error en el voltaje recibido igual a la resolución, 0.01 V. En el segmento 2, para cada dos códigos de 12 bits posibles, hay sólo un código de 12 bits recuperado, por lo tanto se realiza una compresión de codificación de 2:1.

c) Para determinar los códigos para 10.23 V, el proceso es igual:

Código lineal de 12 bits

	No transmitido	
	1	1 1 1 1 1 111111
1	↑	
s		A B C D truncado

Código comprimido de 8 bits

$\begin{array}{c} 1 \\ \uparrow \\ s \end{array}$
 $\begin{array}{c} 111 \\ \text{segmento} \end{array}$
 $\begin{array}{c} 1111 \\ ABCD \end{array}$

Código recuperado de 12 bits

$\begin{array}{c} 1 \\ \uparrow \\ s \end{array}$
 $\begin{array}{c} 1 \\ \uparrow \\ \text{insertado} \end{array}$
 $\begin{array}{c} 1111 \\ \uparrow \\ ABCD \end{array}$
 $\begin{array}{c} 100000 \\ \uparrow \\ \text{insertado} \end{array}$

La diferencia en el código lineal de 12 bits original y el código de 12 bits recuperado es:

$$\begin{array}{r}
 111111111111 \\
 - 111111100000 \\
 \hline
 000000011111 = 31 \text{ (0.01V)} = 0.31 \text{ V}
 \end{array}$$

Para propósitos de comparación, la siguiente fórmula se usa para calcular el porcentaje de error introducido por la compresión digital:

$$\% \text{ de error} = \frac{|\text{voltaje_Tx} - \text{voltaje_Rx}|}{\text{voltaje_Rx}} \quad (\text{IV.36})$$

Ejemplo IV.9

El porcentaje de error máximo ocurrirá para el número más pequeño en el subsegmento más bajo, con cualquier segmento dado. Debido a que no hay error de compresión en los segmentos 0 y 1, para el segmento 3 el máximo porcentaje de error se calcula de la siguiente manera:

Código transmitido de 12 bits: $s00001000000$
 Código recibido de 12 bits: $s00001000010$
 Magnitud de error: 00000000010

$$\% \text{ de error} = \frac{|1000000 - 1000010|}{1000010} \cdot 100 = \frac{|64 - 66|}{66} \cdot 100 = 3.03\%$$

Para el segmento 7:

Código transmitido de 12 bits: $s10000000000$
 Código recibido de 12 bits: $s10000100000$
 Magnitud de error: 00000100000

$$\% \text{ de error} = \frac{|10000000000 - 10000100000|}{1000100000} \cdot 100 = \frac{|1024 - 1056|}{1056} \cdot 100 = 3.03\%$$

Aunque son varias las maneras en las que la compresión de 12 bits a 8 bits y la expansión de 8 bits a 12 bits se pueden lograr con *hardware*, el método más sencillo y económico es con una tabla de consulta en ROM (memoria sólo de lectura).

Fundamentalmente cada función que hace un codificador y decodificador PCM se logra ahora con un solo chip de circuito integrado que se conoce como *codec*. La mayoría de los codecs desarrollados últimamente incluyen un filtro (pasa bandas) de antialiasing, un circuito de muestreo y retención y un convertidor analógico a digital en la sección de transmisión, y un convertidor digital a analógico, un circuito de muestreo y retención y un filtro de pasa bandas en la sección de recepción, que incorpora la corrección del término *sampling*.

IV.5. JERARQUÍAS DE SISTEMAS PCM

La multicanalización es la transmisión de información (ya sea de voz o de datos) de más de una fuente a más de un destino, por el mismo medio de transmisión. Las transmisiones ocurren en el mismo medio, pero no necesariamente al mismo tiempo. El medio de transmisión puede ser un par de cables metálicos, un cable coaxial, un sistema de radio de microondas terrestre, un radio de microondas por satélite, o un cable de fibra óptica. Hay varias formas en que se puede lograr el proceso de multicanalización, aunque los dos métodos más comunes son la multicanalización por división de frecuencia (FDM) y la multicanalización por división de tiempo (TDM).

Con TDM, las transmisiones para fuentes múltiples ocurren sobre el mismo medio pero no al mismo tiempo. Las transmisiones de varias fuentes se intercalan en el dominio del tiempo. El tipo más común de modulación utilizada con los sistemas TDM es PCM. Con un sistema PCM-TDM, se muestran dos o más canales de banda de voz, convertidos a códigos PCM, y luego se utiliza el proceso de multicanalización por división de tiempo en un solo par de cables metálicos o en un cable de fibra óptica.

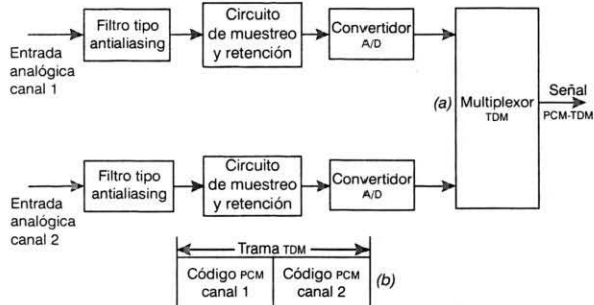


Figura IV.13. Sistema PCM-TDM de dos canales, a) diagrama a bloques, b) trama PCM

La figura IV.13a muestra un diagrama a bloques simplificado de un sistema de portadora para PCM-TDM de dos canales. Cada canal se usa y se convierte de manera alternada a un código PCM. Mientras que el código PCM para el canal 1 se está transmitiendo, el canal 2 se está usando y convirtiendo a un código PCM. Mientras que el código PCM del canal 2 se está transmitiendo, la siguiente señal se toma del canal 1 y se convierte al código PCM. Este proceso continúa y se toman de manera alternativa señales de cada canal, se convierten a códigos PCM y se transmiten.

El proceso de multicanalización simplemente es un interruptor electrónico con dos entradas y una salida. El canal 1 y el canal 2 se seleccionan de manera alterna y se conectan a la salida del multicanalizador. El tiempo que toma transmitir una señal en cada canal se llama tiempo de trama.

El código PCM para cada canal ocupa una ranura de tiempo fija dentro de la trama total de TDM. Con un sistema de dos canales, el tiempo asignado para cada canal es igual a la mitad de la trama del tiempo total. Se toma una señal de cada canal una vez, durante cada trama. Por lo tanto, el tiempo de la trama total es igual al recíproco de la razón de muestreo ($1/f_s$). La figura IV.13b muestra la asignación de la trama TDM para un sistema de dos canales.

IV.5.1. Portadora digital T1

Una portadora digital es un sistema de comunicación que utiliza pulsos digitales para codificar información en lugar de señales analógicas. La figura IV.14 muestra el diagrama a bloques del sistema de portadora digital T1 del sistema Bell. Este sistema es el estándar telefónico en los Estados Unidos. Una portadora T1 multicanaliza por división de tiempo 24 muestras codificadas en PCM para una transmisión, en un solo par de cables metálicos o de fibra óptica. Nuevamente, el multicanalizador simplemente es un interruptor, salvo que ahora tiene 24 entradas y 1 salida. Los 24 canales de banda de voz se seleccionan en secuencia y se conectan a la salida del multicanalizador. Cada canal de banda de voz ocupa un ancho de banda de 300 a 3 400 Hz.

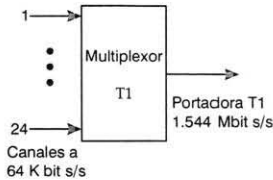


Figura IV.14. Diagrama a bloques de un sistema digital de portadora T1

Simplemente, multicanalizando por división de tiempo 24 canales de banda de voz no constituye en sí una portadora T1. En este momento, la salida del multicanalizador simplemente es una señal digital multicanalizada (DS-1), no se convierte realmente en una portadora T1 hasta que sus líneas estén codificadas y colocadas en los pares de cables, acondicionados especialmente, a las que se conocen como líneas T1. Explicaremos este tema detalladamente más adelante.

Con el sistema de portadora T1 del sistema Bell, los bancos de canales tipo *D* (digital) hacen el muestreo, la codificación y la multicanalización con los 24 canales de banda de voz. Cada canal contiene un código PCM de 8 bits y se muestrea 8 000 veces por segundo. (Cada canal se muestrea a la misma velocidad, pero no necesariamente al mismo tiempo). Por lo tanto, se transmite una señal PCM codificada de 64 Kbits/s por cada canal de banda de voz durante cada trama.

Dentro de cada trama se agrega un bit adicional llamado bit de trama, cuya velocidad es de 8 000 bits/s, se recupera en los circuitos del receptor y se utiliza para mantener la sincronización de la trama y la señal entre el transmisor y el receptor TDM. Como resultado, cada trama TDM contiene 193 bits.

$$\frac{8 \text{ bits}}{\text{canal}} \times \frac{24 \text{ canales}}{\text{trama}} = \frac{192 \text{ bits}}{\text{trama}} + \frac{1 \text{ bit de trama}}{\text{trama}} = \frac{193 \text{ bits}}{\text{trama}}$$

Como resultado la velocidad de la portadora T1 (bps) es:

$$\text{velocidad portadora T1} = \frac{193 \text{ bits}}{\text{trama}} \times \frac{8\,000 \text{ tramas}}{\text{segundo}} = 1.544 \text{ Mbps}$$

Los primeros sistemas de portadora T1 estaban equipados con los bancos de canales D1A que utilizan un código PCM de sólo 7 bits de magnitud con compansión analógica y $\mu = 100$. Una versión posterior del banco de canales D1 (D1D) utiliza códigos PCM, con magnitud de señales de 8 bits. Con los bancos de canales D1A se agrega un octavo bit (el bit *s*) a cada palabra de código PCM, con el propósito de señalización (supervisión, colgado, descolgado, marcar con pulsos, etc.). En consecuencia, la razón de señalización para los bancos de canales D1 es de 8 Kbits/s. Además, con los bancos de canales D1, la secuencia de bits de trama es simplemente un patrón 1/0 alternado. La figura IV.15 muestra la alineación de trama y muestreo para un sistema de portadora T1 utilizando los bancos de canal D1A.

Genéricamente, el sistema de portadora T1 ha progresado a través de los bancos de canales D2, D3, D4, D5 y D6. El D4, D5 y D6 utilizan un código PCM comprimido con magnitud de señales de 8 bits compandido de manera digital con $\mu = 255$. En el banco de canales D1, las características de compresión y expansión se implantaron en los circuitos por separado del codificador y decodificador.

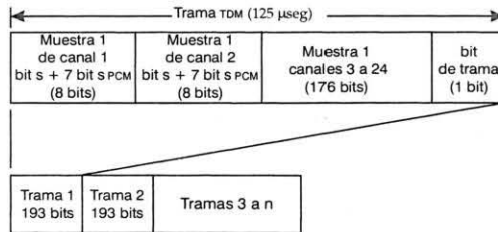


Figura IV.15. Formato de trama T1 utilizando canales D1

Los bancos de canales D2, D3, D4 y D5 incorporan directamente las funciones de compansión en los codificadores y decodificadores. Aunque los bancos de canales D2 y D3 son similares funcionalmente, los bancos de canales D3 fueron los primeros en incorporar, por separado, circuitos integrados LSI personalizados (codecs) para cada canal de banda de voz. Con los bancos de canales D1, D2 y D3, el equipo común realiza las funciones de codificación y decodificación. En consecuencia, un solo mal funcionamiento en el equipo constituye una falla total del sistema.

Los bancos de canales D1A utilizan un código de sólo magnitud; por consiguiente un error en el bit más significativo (MSB), de una muestra de un canal, siempre produce un error decodificado igual a la mitad del rango total de cuantificación ($V_{\text{máximo}}$). Puesto que los bancos de canales D1D, D2, D3, D4 y D5 utilizan un código con magnitud de señales, un error en el MSB (bit de signo) causa un error decodificado igual al doble de la magnitud de la muestra (de $+V$ a $-V$, o viceversa). El peor error es igual al doble del rango total de cuantificación. Sin embargo, las muestras de amplitud máxima ocurren rara vez, y la mayoría de los errores de codificación D1D, D2, D3, D4 y D5 son menores a la mitad del rango de codificación. En promedio, el rendimiento de error con un código de magnitud de señales es mejor que un código de sólo magnitud.

La razón de señalización de 8 Kbit/s que se utiliza con bancos de canales D1 es excesiva para la transmisión de voz. Por lo tanto, con los bancos de canales D2 y D3, un bit de señalización se

sustituye sólo en el bit menos significativo (LSB), cada sexta trama. Por lo tanto, cinco de cada seis tramas tienen resolución de 8 bits, mientras que una de cada seis tramas (la trama de señalización) sólo tiene resolución de 7 bits. En consecuencia, la razón de señalización en cada canal es 1.333 Kbps (8 000 bps/6), y el número efectivo de bits por muestra realmente es $7\frac{5}{6}$ bits y no 8.

Ya que solamente cada sexta trama incluye un bit de señalización, es necesario que todas las tramas estén numeradas para que el receptor sepa cuándo extraer la información de señalización. Además, porque la señalización se efectúa con una palabra binaria de 2 bits, es necesario identificar el MSB y LSB de la palabra de señalización. En consecuencia, se inventó el formato de supertrama que aparece en la figura IV.16. Dentro de cada supertrama, hay 12 tramas numeradas consecutivamente (1-12). Los bits de señalización se sustituyen en las tramas 6 y 12, el MSB en la trama 6 y el LSB en la trama 12. Las tramas 1 a 6 se llaman carretera A, con la trama 6 designada como la trama de señalización de canal A. Las tramas 7 a 12 se llaman carretera B, con la trama 12 designada como la trama de señalización de canal B. Por lo tanto, además de identificar las tramas de señalización, la sexta y la duodécima se deben identificar positivamente.

Para identificar las tramas 6 y 12 se utiliza una secuencia diferente de bits de tramas, en las tramas con números par e impar. Las tramas impares (tramas 1, 3, 5, 7, 9 y 11) tienen un patrón 1/0 alternativo, y las tramas pares (tramas 2, 4, 6, 8, 10 y 12) tienen un patrón repetitivo 0 0 1 1 1 0. Como resultado, el patrón de bits combinados para los bits de trama, es un patrón repetitivo 1 0 0 0 1 1 0 1 1 0 0. Las tramas con números impares se utilizan para la sincronización de tramas y muestras, mientras que las tramas con números pares se utilizan para identificar las tramas de señalización de canales A y B (6 y 12). La trama 6 se identifica por una transición 0/1 en el bit de tramas entre las tramas 4 y 6. La trama 12 se identifica por una transición 1/0 en el bit de tramas, entre las tramas 10 y 12.

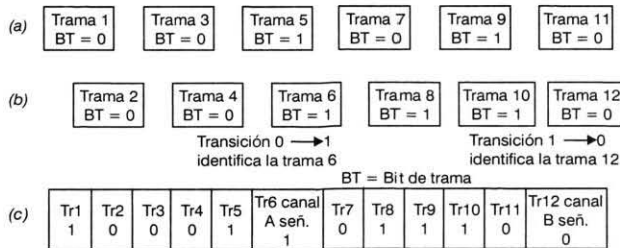


Figura IV.16. Secuencia de bits de tramas o el formato de supertrama T1, utilizando los bancos de canales D2 o D3: a) bits para la sincronización de tramas (tramas con números impares); b) bits de alineación de tramas para la transmisión de señales (tramas con números pares); c) alineación de tramas compuestas

Las tramas para el sistema DS-1 (T1) o el patrón de tramas para los sistemas portadores DS-1C (T1C) de multicanalización por división de tiempo se agregan a la señal digital multicanalizada, a la salida del multicanalizador. La figura IV.17 muestra la circuitería para bits de tramas, para el sistema de portadora T1 de 24 canales, utilizando los bancos de canales (DS-1) D1, D2 o D3. Observe que la razón de bits a la salida del multicanalizador TDM es 1.536 Mbits/s y la razón de bits a la salida del desplazador de registros de 193 bits es de 1.544 Mbits/s. La diferencia (8 Kbits/s) se debe a la adición del bit de tramas en el desplazador de registros.

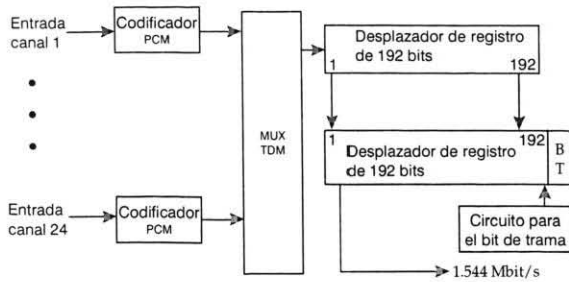


Figura IV.17. Circuitería para el bit de tramas para el sistema de portadora T1 DS1

IV.5.2. Portadora digital E1

La figura IV.18 muestra la alineación de tramas para el sistema PCM-TDM, de estándares europeos de la UIT (Unión Internacional de Telecomunicaciones). Con este sistema, una trama de 125 μ s se divide en 32 ranuras de tiempo iguales. La ranura de tiempo 0 se utiliza para un patrón de alineación de trama y para un canal de alarma. La ranura de tiempo 16 se utiliza para un canal de señalización. La señalización para todos los canales de banda de voz se hace en el mismo canal de señalización; en consecuencia, se utilizan 30 canales de banda de voz en la multicanalización por división de tiempo en cada trama, y por esta característica el sistema se conoce también como PCM 30 + 2.

Con este estándar, cada ranura de tiempo tiene 8 bits, por consiguiente, el número total de bits por trama es de 256, y la velocidad de la portadora E1 es:

$$\text{velocidad portadora E1} = N n f_s = (32) (8) (8\,000) = 2\,048 \text{ Mbits/s}$$

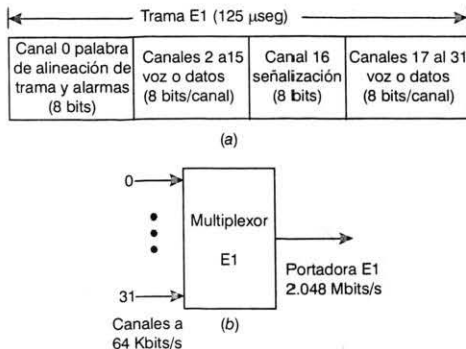


Figura IV.18. a) Formato de trama E1, b) diagrama a bloques de un sistema de portadora E1

En aplicaciones de voz, la portadora E1 debe transmitir información (señalización) para controlar y supervisar cada uno de los canales de voz. La ranura de tiempo 16 es utilizada para este propósito. La señalización se transmite dividiendo los 8 bits del canal 16 en dos bloques de 4 bits

cada uno, de tal manera que los 4 bits menos significativos transportan información de señalización para los canales 17 a 31, y los cuatro bits más significativos sirven para señalar los canales del 1 al 15; la información de estos bits indica, entre otras cosas, si el canal está libre u ocupado, si el teléfono del usuario al que se llama está colgado o descolgado, etc. Debido a que se utiliza el mismo canal para señalar varios canales de voz (un solo canal 16 debe soportar 30 canales de voz), se utiliza una estructura de multitrama para lograr este propósito; en la figura IV.19 se muestra la estructura de multitrama para una vía E1. Para indicar el inicio de la multitrama el canal 16 de la trama 0 transporta la palabra de sincronía de multitrama; en todas las demás tramas el canal 16 se usa como se explicó anteriormente.

El canal cero de cada trama se utiliza para indicar el inicio de trama (palabra de sincronía de trama), así como para la transmisión de alarmas de pérdida de sincronía del enlace; la manera de diferenciar cuando se trata de palabras de sincronía de trama o de alarmas, se hace transmitiendo la palabra de sincronía de trama en la ranura 0 de todas las tramas pares, y la palabra de alarmas se transmite también en la ranura 0 pero sólo en las tramas impares.

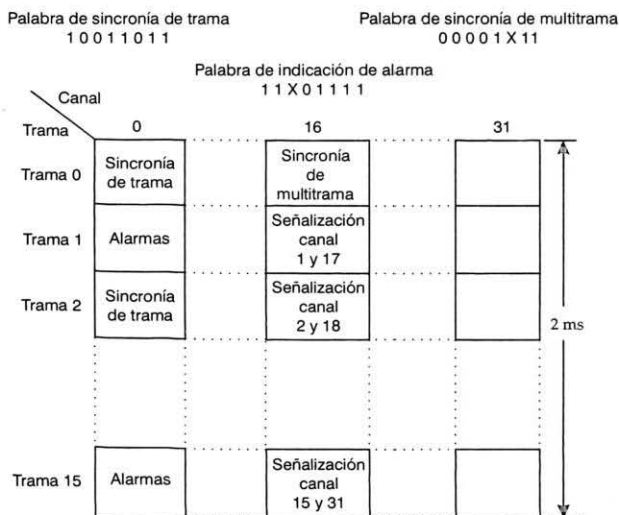


Figura IV.19. Estructura de multitrama utilizada en el enlace E1

Cuando el bit X en la palabra de alarma es puesto a 1 indica que hay una falla en el flujo de datos entrante, esta falla puede deberse a una alta tasa de error, o a la pérdida de alineación de trama. En la palabra de alineación de multitrama cuando se fija a 1, indica que el multiplexor de recepción no puede obtener la alineación de multitrama.

El tipo de señalización que aparece en la figura IV.19 se conoce como señalización asociada al canal, debido a la asignación fija del canal de señalización a cada canal de voz; es decir, siempre se transmite información de señalización para el canal 1 y 17 en la trama 1 de la multitrama; en la trama 2 ahora el canal 16 transporta información pero para los canales 2 y 18, y así sucesivamente hasta la trama 15, donde la señalización corresponde a los canales 15 y 31. En este caso si se utilizan los cuatro bits de señalización asociados a cada canal, la velocidad de señalización por canal es igual a 2 Kbits/s (32 Kbps/16). Una desventaja de este método de señalización es que aunque no haya información

de señalización relevante para el canal (por ejemplo un canal en particular puede pasar mucho tiempo en estado libre —desocupado—, cuando la densidad de tráfico telefónico es baja), debe transmitirse su estado en la trama asociada a él en la estructura de multitrama.

Otro esquema de señalización que utiliza la estructura de trama de la figura IV.18 es el denominado sistema de señalización por canal común No. 7; en este caso se sigue utilizando el canal 16 para transmitir la información de señalización pero ya no se utiliza la estructura de multitrama mostrada en la figura IV.19, es decir no existe una relación fija entre canal de voz y el momento de transmitir su señalización, sino que se asigna por demanda a su capacidad plena (64 Kbits/s); la información de señalización se transmite en forma de paquetes, enviados a través del canal 16, puesto que en este sistema de señalización se soluciona la desventaja de estar transmitiendo información no relevante para los canales de voz y a su alta velocidad de transmisión, entonces es posible señalar varios enlaces E1 con un solo canal de señalización, permitiendo así el poder utilizar los canales de señalización de estos enlaces para la transmisión de voz.

Una comparación de la construcción de la trama y multitrama para el PCM 30 + 2 y el PCM de 24 canales se resume en el cuadro IV.4.

IV.5.3. Multiplexaje digital de alto orden (UIT)

El sistema PCM de 30 canales solamente es el primero de la jerarquía de multiplexaje digital recomendada por la UIT. Si es necesario transmitir más de 30 canales, las jerarquías de más alto orden se construyen como se ilustra en la figura IV.20. Cuatro sistemas de primer orden se combinan (multiplexan) para formar una salida conteniendo 120 canales, a esto se le llama el multiplex de segundo orden (también se conoce como vía o enlace E2). De manera similar, cuatro sistemas de 120 canales pueden multiplexarse para obtener una salida de 480 canales (multiplexaje de tercer orden —vía o enlace E3—); asimismo cuatro sistemas de 480 canales se multiplexan para obtener una salida con 1920 canales (cuarto orden —E4—). Cuatro sistemas de 1920 canales son a su vez multiplexados para obtener 7 680 canales (quinto orden —E5—) a la salida. Éste es el mayor nivel de multiplexaje hoy en día.

CUADRO IV.4. Comparación de los sistemas de 24 y 30 canales

	PCM de 24 canales	PCM 30 + 2
Frecuencia de muestreo (kHz)	8	8
Duración de la ranura de tiempo (μs)	5.2	3.9
Anchura de bit (μs)	0.65	0.49
Razón de transferencia de bit (Mbit/s)	1.544	2.048
Periodo de trama (μs)	125	125
Número de bits por palabra	8	8
Número de tramas por multitrama	12	16
Periodo de multitrama (ms)	1.5	2
Señal de alineamiento de trama en	Tramas impares	Tramas pares
Señal de alineamiento de multitrama en	Tramas pares	Ranura 16 Trama 0
Palabra de alineamiento de trama	101010	10011011
Palabra de alineamiento de multitrama	001110	00001011

La salida a 8 Mb/s de un multiplexor de segundo orden se forma multiplexando cuatro salidas multiplex de primer orden. Esto se hace intercalando los flujos de bits de los cuatro sistemas primarios. El principal problema que se debe resolver en este proceso es la organización de los cuatro tributarios. Existen dos categorías de multiplexaje digital.

- Multiplexores digitales síncronos
- Multiplexores digitales asíncronos

Los multiplexores digitales síncronos tienen tributarios con el mismo reloj y frecuencia, y todos ellos están sincronizados a un reloj maestro. Los multiplexores digitales asíncronos, por su parte, tienen tributarios con la misma frecuencia nominal (esto significa que puede haber una pequeña diferencia entre uno y otro), pero no están sincronizados entre sí. La diferencia entre estos dos tipos de multiplexaje se torna aparente cuando se imagina la situación donde los cuatro tributarios se mezclan. Para el caso síncrono, los pulsos de cada tributario tienen sus transiciones de subida y bajada en los mismos instantes de tiempo. Para el caso asíncrono los tiempos de subida y bajada de los pulsos no coinciden unos con otros. El diseño asíncrono se usa con más frecuencia en la actualidad, pero los sistemas síncronos serán los que predominen en un futuro cercano. La técnica de multiplexaje la explicaremos a continuación. El multiplexaje de varias fuentes tributarias puede hacerse por:

- Intercalado/multiplexaje bit a bit
- Intercalado/multiplexaje palabra por palabra

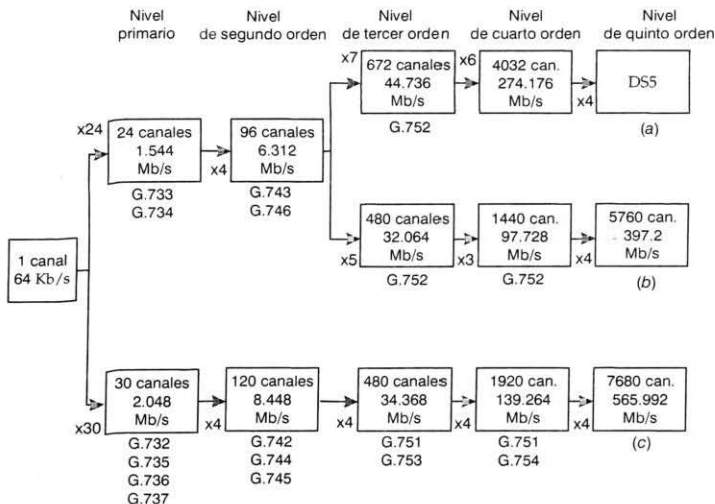


Figura IV.20. La jerarquía digital asíncrona. a) Norteamérica, b) Japón y c) Europa

Las figuras IV.21a y IV.21b ilustran la diferencia entre intercalado de bits e intercalado de palabras. Los términos se explican por sí mismos. Por ejemplo, en la figura IV.21a hay cuatro flujos de bits (tributarios) que van a ser multiplexados. Se toma un bit secuencialmente de cada tributario,

de tal manera que el flujo de bits multiplexado que resulte tendrá cada cuatro bits un bit de información del mismo tributario. En la figura IV.21b se toma un número específico de bits, para formar una palabra de cada tributario en turno.

El intercalado de palabras requiere algunas restricciones sobre la estructura de trama de los tributarios y una mayor capacidad de memoria en el multiplexor. El intercalado bit a bit es mucho más simple porque es independiente de la estructura de trama de los tributarios y requiere mucho menos capacidad de memoria; por lo tanto, generalmente se utiliza en sistemas de multiplexaje asíncrono.

El multiplex de segundo orden (E2) tiene cuatro tributarios de primer orden cada uno a una velocidad de 2.048 Mbit/s, intercalados bit a bit; cabe hacer notar que la salida del MUX E2 de 8.448 Mbit/s no es exactamente cuatro veces la velocidad de 2.048 Mbit/s. Éste es un resultado de la naturaleza asíncrona del sistema, que estudiaremos a continuación. Cada tributario tiene su propio reloj y está temporizado con lo que suele llamarse una frecuencia pleisocrónica, esto es, una frecuencia nominal a la cual se le permite tener cierta tolerancia (deriva) fijada dentro de ciertos límites; por ejemplo, la salida del multiplexor primario es de 2.048 Mbit/s ± 50 ppm (pulsos por millón). Entonces, para soportar estas pequeñas variaciones de las frecuencias de los tributarios alrededor de su valor nominal, cuando se multiplexan cuatro tributarios al siguiente nivel de la jerarquía, se utiliza un proceso conocido como relleno positivo (*pulse stuffing*) o justificación positiva.

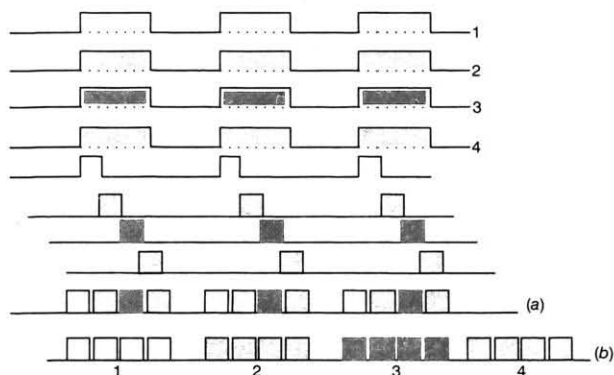


Figura IV.21. a) Multiplexaje bit a bit, b) multiplexaje palabra por palabra

La justificación positiva implica el uso de un canal de salida cuya razón de bit es a propósito mayor que la razón de bit de entrada, de ahí que el canal de salida contenga todos los bits de la señal de entrada más un número variable de *bits de relleno* que no forman parte de la información de los canales de entrada. Estos bits de relleno se insertan en posiciones específicas, lo que provoca que el flujo de salida tenga una mayor razón de bits, por lo que deben identificarse en el receptor y por lo tanto quedan descartados para poder recuperar el flujo de bits original. La justificación positiva se utiliza en multiplexaje de alto orden cuando cada uno de los tributarios de menor nivel jerárquico no están sincronizados, ni existe por lo tanto una relación de fase prefijada entre ellos. Esta situación se observa en las figuras IV.22a y IV.22b. La razón de bit de las entradas en la figura IV.22a son exactamente las mismas y los pulsos arriban de manera sincronizada. Subsecuentemente la salida tiene un intercalado de bit perfecto; sin embargo, en la figura IV.22b, las razones de bits de las entradas no son idénticas, es decir los pulsos no arriban de manera sincronizada. La diferencia en

las razones de bits está exagerada para propósitos del ejemplo. Ésta es la situación que aparece para un sistema de multiplexaje asíncrono, donde la velocidad de los tributarios no son exactamente las mismas; de ahí que la salida del sistema en la figura IV.22b requiere algunos pulsos de justificación adicionales para soportar las diferencias en las razones de bits de las entradas. Aunque el método utilizado en multiplexaje asíncrono es el de intercalado de bits, para propósitos de comparación con los sistemas síncronos en la figura IV.22b se muestra una situación de intercalado de palabras. Cada palabra en la señal de salida de la figura IV.22b contiene un número fijo de ranuras de tiempo, asimismo cada palabra contiene un bit de control de relleno bit C y un bit de relleno bit S. Cuando un bit de control se pone a 0, el bit S contiene información de los datos; cuando el bit C se pone a 1, el bit S respectivo es un bit de relleno y se fija a 1.

La estructura de trama para el multiplexor de segundo orden (E2) aparece en la figura IV.23. La trama contiene:

- La palabra de alineación de trama
- Los bits de control de relleno
- Los bits de relleno
- Los bits de los tributarios

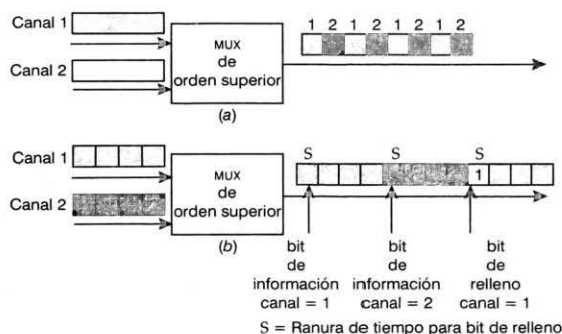


Figura IV.22. Ejemplo simplificado de justificación positiva, a) las razones de bits de las entradas son iguales, b) las razones no son iguales

Es importante hacer notar que la estructura de trama de orden superior no está relacionada con la forma que tenga la estructura de trama del nivel inferior de los tributarios; de hecho, para el multiplexor de orden superior, cada tributario se trata simplemente como un flujo de bits; todos los bits de este flujo, incluyendo los bits de alineación de trama, los bits de alineación de multitrama, los bits de señalización así como los bits de información (canales de voz, datos o video), se transmiten y se tratan de la misma manera por el multiplexor.

Subtrama	Señal	Número de bits	Numeración de bits
1	Palabra de alineación de trama 1111010000	10	1 a 10
	Bits de servicio	2	11 a 12
	Bits de los tributarios	200	13 a 212
2	Bits de control de relleno	4	213 a 216
	Bits de los tributarios	208	217 a 424
3	Bits de control de relleno	4	425 a 428
	Bits de los tributarios	208	429 a 636
4	Bits de control de relleno	4	637 a 640
	Bits de relleno	4	641 a 644
	Bits de los tributarios	204	645 a 848

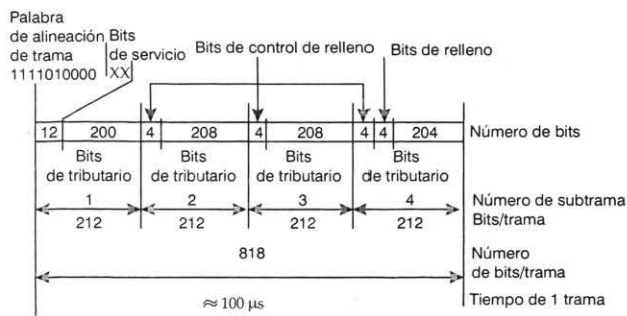


Figura IV.23. Estructura de trama a 8.448 Mb/s (justificación positiva)

La trama del MUX E2 contiene 848 bits y se divide en cuatro subtramas, cada una con 212 bits. Los primeros 12 bits en cada trama contienen 10 bits de alineación de trama (1111010000) y 2 bits de servicio que se utilizan para la transmisión de alarmas, errores de sincronía, etc. Se utiliza una palabra de control de relleno de 3 bits para cada tributario, ocupando un total de 12 bits (4 bits después de cada subtrama). Finalmente, los bits de relleno para cada tributario se insertan (si es requerido) antes del inicio de la cuarta subtrama. La señal de cada tributario se escribe en una memoria elástica utilizando como reloj de escritura la señal de 2.048 Mb/s, extraída de la señal entrante. Una memoria elástica expande la anchura del bit para permitir que la lectura tome lugar durante un intervalo de tiempo mayor (figura IV.24). El dato se lee y transmite por el multiplexor a una frecuencia de reloj de 2.112 Mb/s. El proceso de escritura es inhibido en las ranuras de tiempo específicas para incluir los bits de alineación de trama, los bits de servicio y los bits de control de relleno. Cuando el control de justificación positiva del mensaje requiere un bit de control de relleno, éste se genera cuando las fases de lectura y escritura de la memoria alcanzan un nivel predeterminado. La memoria elástica se efectúa simplemente por medio de un circuito de flip-flop que mantiene el estado 1 hasta el siguiente ciclo de reloj.

Cuando un bit de relleno se inserta, los bits de control de relleno lo indican. Esto consiste en fijar los bits de control a 111 (es decir un 1 en cada uno de los tres espacios asignados a los bits de

control antes de cada subtrama). Estos tres conjuntos de cuatro bits forman las palabras de control de relleno para cada uno de los cuatro tributarios; cuando el relleno no toma lugar, la palabra de control se transmite como 000. Este procedimiento se ilustra en la figura IV.25, donde el primer subíndice del bit C denota el número del tributario, y el segundo subíndice denota uno de los tres bits de la palabra de control para cada tributario. En la recepción, si se detecta una mayoría de unos (dos de tres) en la palabra de control, el bit de justificación se cancela anulando la recepción en la memoria elástica; es decir, en este caso el bit de relleno no forma parte de la información del tributario. A la inversa, si hay una mayoría de ceros (dos de tres) en la palabra de control en el receptor, el bit de relleno no se cancela, es decir, forma parte de los bits de información del tributario correspondiente.

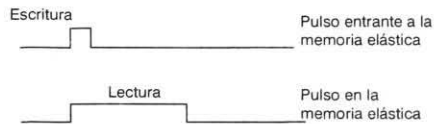


Figura IV.24. Pulsos de escritura y lectura en la memoria elástica

El reloj de escritura en la memoria elástica del receptor tiene las mismas características que el reloj de escritura del transmisor; esto es, tiene una frecuencia en promedio igual que la del tributario, pero con espacios periódicos para la estructura de trama y espacios aleatorios para el proceso de relleno. La memoria elástica se lee con un reloj de 2.048 Mbits/s amarrado en fase al reloj de escritura por medio de un circuito de lazo sujetador de fase (PLL), este circuito permite reducir el jitter causado por la estructura de trama, los componentes de alta frecuencia de jitter provocadas por el proceso de justificación positiva, el jitter provocado por los tributarios y el jitter introducido por el enlace a 8.448 Mbits/s. La estructura de trama para los niveles jerárquicos mayores al de 8.448 Mbits por segundo es similar a la de segundo orden (véase por ejemplo Winch [1]).

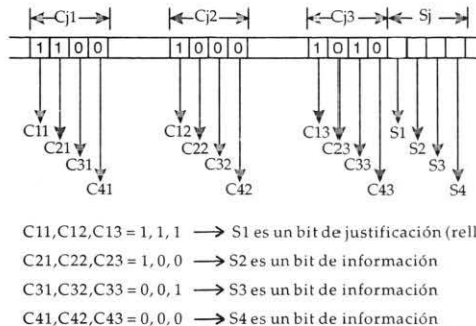


Figura IV.25 Formación de la palabra de control de justificación

IV.6. CODIFICACIÓN DE CANAL

La codificación de canal involucra convertir los niveles lógicos estándar (TTL, CMOS y otros parecidos) a una forma más adecuada para la transmisión sobre líneas telefónicas. Fundamentalmente deben considerarse cinco factores principales para seleccionar un formato para la codificación de canal:

1. Voltajes de transmisión y componentes de cd
2. Recuperación de temporización (reloj)
3. Ancho de banda de transmisión
4. Facilidad de detección y decodificación
5. Detección de errores

Voltajes de transmisión y componentes de cd

Los voltajes o niveles de transmisión pueden entrar en dos categorías: en la unipolar (UP) o en la bipolar (BP). La transmisión unipolar de datos binarios involucra la transmisión solamente de un nivel de voltaje, que no es cero (por ejemplo, + V para 1 lógico y 0 V o tierra para 0 lógico). En la transmisión bipolar están involucrados dos niveles de voltaje que no son cero (por ejemplo, + V para 1 lógico y -V para 0 lógico).

En una línea de transmisión digital se usa con más eficiencia la potencia, codificando datos binarios con voltajes que son iguales en magnitud, pero opuestos en polaridad y balanceados simétricamente, cerca de 0 V. Por ejemplo, suponiendo una resistencia de $1\ \Omega$ y un nivel de 1 lógico de +5 V y un nivel de 0 lógico de 0 V, la potencia promedio requerida es 12.5 W (suponiendo que sea igualmente probable la ocurrencia de un 1 o 0). Con un nivel de 1 lógico de +2.5 V y un nivel de 0 lógico de -2.5 V, la potencia promedio es sólo 6.25 W. Por lo tanto, al utilizar voltajes simétricos bipolares, la potencia promedio se reduce por un factor de 50 por ciento.

Ciclo de trabajo

El ciclo de trabajo de un pulso binario puede utilizarse también para ordenar el tipo de transmisión. Si el pulso binario se mantiene por el total del tiempo de bit, se llama no regreso a cero (NRZ); si el tiempo activo del pulso binario es menor que 100% del tiempo de bit, se llama regreso a cero (RZ).

Los voltajes de transmisión unipolar y bipolar, así como la codificación de regreso a cero y no regreso a cero pueden combinarse de varias formas para alcanzar un esquema de codificación de línea en particular; la figura IV.26 muestra cinco posibilidades para codificación de canal.

En la figura IV.26a solamente hay un nivel de voltaje que no es cero (+V = 1 lógico); un voltaje de cero simplemente implica un 0 binario. Además, cada 1 lógico mantiene el voltaje positivo para todo el tiempo de bit (100% del ciclo de trabajo); en consecuencia, la figura IV.26a representa una señal no regreso a cero unipolar (UPNRZ). En la figura IV.26b hay dos voltajes que no son cero (+V = 1 lógico y -V = 0 lógico) y se utiliza un ciclo de trabajo de 100%. La figura IV.26b representa una señal no regreso a cero bipolar (BPNRZ). En la figura IV.26c solamente se utiliza un voltaje que no es cero, pero cada pulso está activo solamente 50% del tiempo de bits. En consecuencia, la figura IV.26c representa una señal regreso a cero unipolar (UPRZ). En la figura IV.26d hay dos voltajes que no son cero (+V = 1 lógico y -V = 0 lógico). Además, cada pulso está activo solamente 50% del total

del tiempo de bits, por consiguiente, la figura IV.26d representa una señal regreso a cero bipolar (BPRZ). En la figura IV.26e nuevamente hay dos niveles de voltaje que no son cero ($-V$ y $+V$), pero aquí ambas polaridades representan un 1 lógico y 0 V representa un 0 lógico. Este método de codificación se llama inversión de marca alterna (AMI). Con transmisiones AMI, cada 1 lógico sucesivo se invierte en polaridad del 1 lógico anterior, debido a que se utiliza el regreso a cero; esta técnica de codificación se llama inversión de marca alterna bipolar de regreso a cero (BPRZ-AMI).

Con la codificación NRZ, una cadena larga de unos o ceros produce una condición en donde un receptor puede perder su referencia de amplitud por discriminación óptima entre unos y ceros recibidos. Esta condición se llama desviación de cd . El problema puede surgir también cuando hay un desequilibrio significativo en el número de unos y ceros transmitidos. La figura IV.27 muestra cómo la desviación de cd se produce por una larga cadena de unos lógicos sucesivos. En esta figura puede verse que después de una larga cadena de unos, los errores de 1 a 0 son más probables que los errores de 0 a 1; en forma similar, largas cadenas de ceros incrementan la probabilidad de un error de 0 a 1.

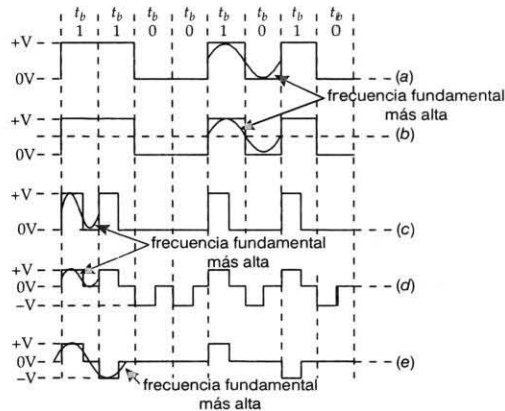


Figura IV.26. Formatos de codificación de línea: a) UPNRZ, b) BPNRZ, c) UPRZ, d) BPRZ y e) BPRZ-AMI

El método utilizado para codificar en línea determina el ancho de banda mínimo que se requiere para transmitir, qué tan fácilmente se puede extraer un reloj de él, qué tan fácilmente se puede decodificar el nivel de cd promedio y si ofrece una forma conveniente de detectar errores.

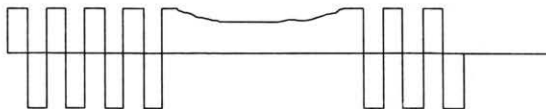


Figura IV.27. Desviación de cd

Consideraciones de ancho de banda

Para determinar el ancho de banda mínimo requerido para propagar una señal codificada en línea se tiene que determinar la frecuencia fundamental más alta asociada con ella (véase la figura IV.26). La frecuencia fundamental más alta se determina con el peor caso (transición más rápida) de secuencia de bits binarios. Con UPNRZ la condición del peor caso es una secuencia 1/0 alternativa; el periodo de la frecuencia fundamental más alta toma el tiempo de 2 bits y, por lo tanto, es igual a la mitad de la tasa de bits. Con BPNRZ, nuevamente la condición del peor caso es una secuencia 1/0 alterna, y la frecuencia fundamental más alta es la mitad de la tasa de bits. Con UPRZ, la condición del peor caso son dos unos sucesivos. El ancho de banda mínimo es por lo tanto igual a la tasa de bits. Con BPRZ, la condición del peor caso son 0 unos o 0 ceros sucesivos, y el ancho de banda mínimo es nuevamente igual a la tasa de bits. Con BPRZ-AMI, la condición del peor caso son dos o más unos consecutivos, y el ancho de banda mínimo es igual a la mitad de la tasa de bits.

Recuperación del reloj

Para recuperar y mantener la información del reloj de los datos recibidos, debe haber un número suficiente de transiciones en la señal de datos. Con UPNRZ y BPNRZ, una cadena larga de unos o ceros consecutivos genera una señal de datos libre de transiciones y, por lo tanto, inadecuada para la sincronización de relojes. Con UPRZ y BPRZ-AMI, una larga cadena de ceros también genera una señal de datos libre de transiciones. Con BPRZ, una transición ocurre en cada posición de bit sin importar si el bit es un 1 o un 0. En el circuito de recuperación del reloj, los datos simplemente se rectifican a onda completa para producir un reloj independiente de los datos igual a la tasa de bits de recepción. Por consiguiente, la codificación BPRZ es más adecuada para la recuperación del reloj. Si se evita que ocurran largas secuencias de ceros, la codificación BPRZ-AMI es suficiente para asegurar la sincronización del reloj.

Detección de errores

Con las transmisiones UPNRZ, BPNRZ, UPRZ y BPRZ, no hay forma de determinar si los datos recibidos tienen errores. Con las transmisiones BPRZ-AMI, un error en cualquier bit causará una violación bipolar (la recepción de dos o más unos consecutivos, con la misma polaridad). Por lo tanto, BPRS-AMI tiene un mecanismo integrado para detectar errores.

Facilidad de detección y decodificación

Debido a que la transmisión unipolar involucra la transmisión de sólo un voltaje de polaridad, un voltaje promedio de cd se asocia con la señal igual a $+V/2$. Suponiendo que exista igual probabilidad de que ocurran unos y ceros, las transmisiones bipolares tienen un componente promedio de cd de 0V. Un componente de cd es poco deseable, ya que polariza la entrada a un detector de umbral convencional (un comparador polarizado) y podría causar una mala interpretación de la condición lógica de los pulsos recibidos. Por lo tanto, la transmisión bipolar es más adecuada para la detección de datos.

El cuadro IV.5 resume el ancho de banda mínimo, el voltaje promedio cd , la recuperación del reloj, y las capacidades de detección de errores, de los formatos, para la codificación en línea

mostrados en la figura IV.26. En el cuadro IV.5 puede verse que la codificación BPRZ-AMI tiene las mejores características en general; por consiguiente, es el método que se utiliza más comúnmente.

CUADRO IV.5. Resumen de códigos de línea. (*) Denota el mejor funcionamiento o calidad

Formato de codificación	BW mínimo	DC promedio	Recuperación de reloj	Detección de errores
UPNRZ	$f_b/2^*$	$+V/2$	Malo	No
BPNRZ	$f_b/2^*$	$0V^*$	Malo	No
UPRZ	f_b	$+V/2$	Bueno	No
BPRZ	f_b	$0V^*$	El mejor*	No
BPRZ-AMI	$f_b/2^*$	$0V^*$	Bueno	Si*

Bifase digital

La bifase digital (a veces llamada código de Manchester o difase) es un tipo popular de codificación en línea que produce un componente de tiempo fuerte para la recuperación del reloj y no causa la desviación de cd . La bifase es una forma de transmisión BPRZ que utiliza un ciclo de onda cuadrada en fase de 0° para representar un 1 lógico y un ciclo de onda cuadrada en fase de 180° para representar un 0 lógico. La codificación de bifase digital aparece en la figura IV.28. Observe que ocurre una transición en el centro de cada elemento, sin importar cuál sea su fase. Por lo tanto la bifase produce un fuerte componente de sincronización para la recuperación del reloj. Además, suponiendo una probabilidad igual de unos y ceros, el voltaje cd promedio es cero volts y no hay una desviación de cd . Una desventaja de la bifase es que no contiene ninguna forma de detección de errores.

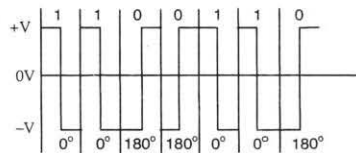


Figura IV.28. Código bifase digital

Código HDB3

El objetivo del código HDB3 es limitar el número de ceros que puedan aparecer en forma consecutiva, a solamente tres, lo que asegura extracción de temporización en el regenerador del receptor. Este código ha sido recomendado por la UIT (G.703) para utilizarse en los sistemas a 2, 8 y 34 Mbits/s. En la figura IV.29 vemos un ejemplo de este código. Secuencias mayores a tres ceros se evitan reemplazándolas por secuencias que sólo tienen uno o dos ceros de acuerdo con ciertas reglas específicas. Estas sustituciones se detectan porque producen una violación al código AMI; además, las secuencias de reemplazo no introducen una componente de dc apreciable. Las reglas del código HDB3 pueden resumirse como sigue:

1. Aplique las tres reglas etapa por etapa.

2. Generación de sustituciones 000V y B00V:

a) Se sustituye 000V si hay un número par de unos desde el último pulso de violación. El pulso V es un pulso que viola la regla del código AMI:

b) Se sustituye B00V si hay un número impar de unos desde el último pulso de violación. El pulso B es un pulso que debe seguir para su codificación la regla del código AMI.

3. Se deben observar las polaridades del pulso de datos y violación precedentes.

Las anteriores reglas se resumen en el cuadro VI.6.

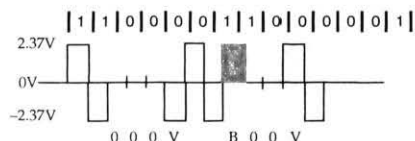


Figura IV.29. Ejemplo de codificación HDB3

CUADRO IV.6. Reglas de sustitución del código HDB3

Polaridad del último pulso V	Polaridad del dato precedente	Secuencia de sustitución
+	+	-00- B00V
-	-	+00+ B00V
+	-	000- 000V
-	+	000+ 000V

Códigos de sustitución binaria de N ceros BNZS

Los códigos BNZS se utilizan en Norteamérica. Estos códigos sustituyen una cadena de N-ceros por una palabra de longitud N, la cual produce una violación bipolar. En el receptor los datos originales se obtienen reconociendo las violaciones bipolares y reemplazándolas por la cadena original de N-ceros. El código de línea B3ZS se especifica en la recomendación G.703 de la UIT, para utilizarse en la interface a 44.736 Mbits/s. La sustitución binaria de tres ceros (B3ZS) es muy similar al código HDB3. En la figura IV.30a se muestra la regla de codificación para B3ZS y en IV.30b un ejemplo de codificación.

(a)	Polaridad del pulso precedente	Número de pulsos bipolares después de la última sustitución	
		Par	Impar
	-	00-	+0+
	+	00+	-0-

(b)

1 1 1 0 0 0 1 1 0 0 0 0 0 0 0 1 0 0 0 1

+ 0 - 0 0 - + - + 0 + - 0 - 0 0 + 0 0 + -

Figura IV.30. Código B3ZS, a) regla de codificación, b) ejemplo de codificación, los signos +, 0 y -, representan pulsos +V, 0V y -V respectivamente

Otro código de esta familia utilizado en Norteamérica es el código B6ZS, empleado por la Bell en sus sistemas a 6.312 Mbit/s, utilizando como medio de transmisión par trenzado. Las violaciones bipolares se introducen en la segunda y quinta posiciones de bit en la palabra de sustitución, es decir cada bloque de seis ceros consecutivos se reemplaza por 0VB0VB, tal como se muestra en el cuadro IV.7.

CUADRO IV.7. Reglas de sustitución del código B6ZS

Polaridad del pulso precedente	Secuencia de sustitución
-	0 - + 0 + -
+	0 + - 0 - +

El código de línea B8ZS está especificado en la recomendación G.703 de la UIT para utilizarse en el multiplexor a 6.312 Mbit/s usado como medio de transmisión cable coaxial. En este código, cada bloque de ocho ceros sucesivos es reemplazado por la secuencia 000VB0VB. En el cuadro IV.8 se presenta un resumen de los principales códigos de línea utilizados por diversas administraciones telefónicas.

CUADRO IV.8. Características principales de códigos de línea

Nombre de la portadora	País o administración	Velocidad (Mbit/s)	Código	Medio de transmisión	Distancia entre repetidores
T1 Primario (E1)	ATT	1.544	Bipolar	Par trenzado	6 000 pies
	UIT	2.048	HDB3	Par trenzado	2 000 metros
	ATT	3.152	Bipolar	Par trenzado	6 000 pies
T1C	ATT	3.152	Bipolar	Par trenzado	6 000 pies
T148	ITT	2.37 (ternario)	4B3T	Par trenzado	6 000 pies
T2	ATT	6.312	B6ZS	Par trenzado de baja capacitancia	14 800 pies
LD-4	Canadá	274.176	B3ZS	Coaxial	1 900 metros
T4M	ATT	274.176	Polar binario (NRZ)	Coaxial	5 700 pies

IV.7. PROBABILIDAD DE ERROR EN TRANSMISIÓN BINARIA

En esta sección analizaremos el efecto que el ruido produce en la calidad de transmisión en un sistema de comunicación digital. Para ser más precisos, supongamos que durante la transmisión se les ha sumado ruido a las señales transmitidas, de tal manera que el voltaje medido en el receptor consista de la suma de los dos, tal como se ve en la figura IV.31.

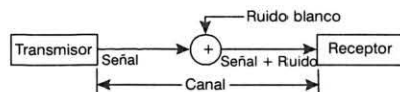


Figura IV.31. Diagrama simplificado de un sistema de comunicación con ruido blanco aditivo

Supongamos que el ruido en el canal es blanco, con varianza igual a σ^2 , media cero y densidad de potencia constante en todo el espectro; este tipo de ruido es el más usado en modelos estadísticos para el ruido aditivo en sistemas de comunicaciones; su función de densidad de probabilidades es del tipo gaussiano; por lo que a veces también se le conoce como ruido gaussiano, su *fdp* está dada por

$$f(x) = \frac{e^{-x^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} \quad (\text{IV.37})$$

en la figura IV.32 vemos su función de densidad de probabilidades.

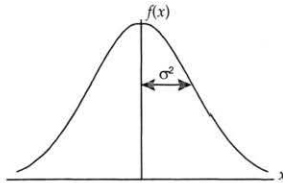


Figura IV.32. Función de densidad de probabilidad del ruido blanco

Ahora supongamos que se reciben pulsos binarios en un sistema de comunicaciones digitales. El ruido $n(t)$ se suma a los pulsos en el receptor, por lo que existe la posibilidad de que pueda causar un error en el proceso de decodificación del receptor, específicamente si se transmiten pulsos codificados en NRZ, donde la presencia de pulso representa los unos del código binario y la ausencia de pulso representa los ceros; el error puede ocurrir por ejemplo si el ruido en la ausencia de pulso tiene una amplitud instantánea comparable a cuando está un pulso presente, o si el ruido en presencia del pulso tiene una amplitud negativa que puede destruir al pulso. En el primer caso el ruido puede confundirse con la presencia de pulso y un 0 puede recuperarse como un 1; en el segundo caso el 1 realmente transmitido puede aparecer como un 0 a la salida del decodificador.

Entonces, nos interesa poder contestar las siguientes preguntas: ¿qué tan frecuentemente pueden ocurrir estos errores en promedio?, ¿es posible reducir el número de errores por debajo de un número tolerable incrementando la amplitud del pulso?, ¿si es así, cuál debe ser el incremento necesario?; ¿cómo se afecta la razón de error al disminuir el ruido? Todas estas preguntas se resuelven conociendo la estadística del ruido o teniendo un modelo razonablemente bueno para él. Contestaremos estas preguntas usando el modelo de ruido de la ecuación (IV.37).

Supongamos que la amplitud de todos los pulsos es de A volts. La señal en el receptor (ruido + información) se muestrea en cada intervalo de bit y cada muestra se compara con un nivel de decisión para determinar si un 1 o un 0 está presente. Este proceso de muestreo ya lo explicamos en el capítulo III, en la sección de interferencia entre símbolos, donde se dedujo cuáles deberían ser los instantes de muestreo óptimos para evitar ISI; aquí estamos interesados en determinar cuál debe ser el nivel de decisión óptimo. Supongamos que el nivel de decisión es igual a $A/2$; entonces si la muestra es menor a $A/2$, se decide que hay un 0 presente; en caso contrario habrá un 1 presente. Los errores pueden ocurrir si, cuando un 1 está presente, el voltaje de la muestra (señal más ruido) es menor que $A/2$, o cuando el pulso está ausente si el ruido excede el nivel $A/2$. Un ejemplo que ilustra los dos tipos de errores es el de la figura IV.33, donde se supone que las muestras se toman al centro de cada intervalo de bit por cuestiones de claridad.

Transmisión digital en banda base

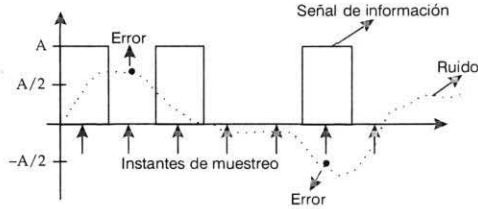


Figura IV.33. Efectos del ruido en transmisión de pulsos binarios

Para determinar la probabilidad de error cuantitativamente, consideremos los dos tipos de error de forma separada: supongamos primero que se transmitió un cero, de tal manera que el pulso no está presente en el decodificador. La probabilidad de error en este caso justamente es la probabilidad de que el ruido exceda en amplitud el nivel de decisión $A/2$. Alternativamente, como $v(t) = n(t)$ si un cero está presente, el valor de la muestra v es una variable aleatoria con la misma estadística que el ruido. Entonces la probabilidad de error es precisamente la probabilidad de que v tome algún valor entre $A/2$ e ∞ . Entonces la función de densidad de probabilidad para v suponiendo un cero presente es

$$f_0(v) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-v^2/2\sigma^2} \quad (\text{IV.38})$$

El subíndice 0 denota la presencia de un símbolo 0 y la probabilidad de error P_{e0} es el área bajo la curva de $f_0(v)$ de $A/2$ a ∞ . La función de densidad $f_0(v)$ aparece en la figura IV.34, donde la probabilidad de error corresponde al área sombreada.

$$P_{e0} = \text{Prob} \left(v > \frac{A}{2} \right) = \int_{A/2}^{\infty} f_0(v) dv \quad (\text{IV.39})$$

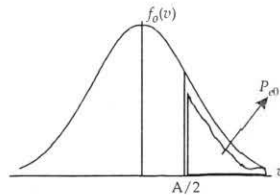


Figura IV.34. Densidad de probabilidad cuando se transmite un cero

Supongamos ahora que transmitimos un 1, éste aparece al decodificador como un pulso de amplitud A más el ruido superpuesto. Una muestra $v(t)$ de la señal compuesta que se toma al tiempo t ahora es la variable aleatoria $A + n(t)$. La cantidad fija A desplaza el nivel de ruido de un promedio de cero volts a un promedio de A volts, por lo que la variable aleatoria v tiene la misma estadística que $n(t)$, pero ahora fluctuando alrededor de A . Esta función de densidad es la misma función de densidad gaussiana con la misma varianza, pero con un valor promedio A ; es decir

$$f_1(v) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(v-A)^2/2\sigma^2} \quad (\text{IV.40})$$

función de densidad tal como se muestra en la figura IV.35.

La probabilidad de error en este caso corresponde a la probabilidad de que la muestra v de señal más ruido esté por debajo de $A/2$; ésta justamente es el área bajo la curva de $f_1(v)$ de $-\infty$ a $A/2$, dada por

$$P_{e1} = \text{Prob} \left(v < \frac{A}{2} \right) = \int_{-\infty}^{A/2} f_1(v) dv \quad (\text{IV.41})$$

esta probabilidad de error se indica por el área bajo la curva de la figura IV.35.

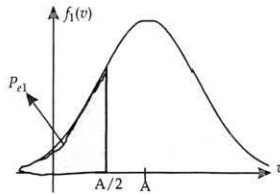


Figura IV.35. Densidad de probabilidad cuando se transmite un uno

Determinaremos ahora la probabilidad de error total del sistema: primero nótese que los dos tipos de errores posibles pueden considerarse eventos mutuamente exclusivos: la aparición de un 0 evita la aparición de un 1 y viceversa; luego las probabilidades pueden sumarse. Sin embargo, en este caso es aparente que P_{e0} y P_{e1} son probabilidades condicionales; la primera se obtiene suponiendo un cero presente y la segunda un 1 presente. Para remover esta condicionalidad debemos multiplicar cada una por su propia probabilidad de ocurrencia. Supongamos que la probabilidad de transmitir un cero la conocemos y la denotamos por P_0 ; asimismo, que la probabilidad de transmitir un 1 es conocida y la identificamos como P_1 , por lo que la probabilidad de error total del sistema es:

$$P_e = P_0 P_{e0} + P_1 P_{e1} \quad (\text{IV.42})$$

Es aparente a partir de la ecuación anterior y tomando en cuenta la simetría de la función gaussiana, que las dos probabilidades condicionales P_{e0} y P_{e1} son iguales, además suponemos que los dos símbolos binarios 0 y 1 son equiprobables, es decir $P_0 = P_1 = 1/2$; encontramos que la probabilidad de error total P_e es la misma que P_{e0} o P_{e1} , es decir

$$P_e = P_{e0} = P_{e1} \quad (\text{IV.43})$$

Recordando las funciones de error $\text{Erf}(x)$ y de error complementario $\text{Erfc}(x)$, tabuladas en diversos libros de estadística y en tablas matemáticas, ambas definidas por

$$\text{Erf}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy \quad (\text{IV.44})$$

$$\text{Erfc}(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-y^2/2} dy \quad (\text{IV.45})$$

entonces, tan sólo haciendo un simple cambio de variable en las ecuaciones (IV.39) y (IV.41) que se deja como un ejercicio para el lector, obtenemos que la probabilidad de error total es:

$$P_e = \text{Erfc}(A/2\sigma) = 1/2 [1 - \text{Erf}(A/2\sqrt{2}\sigma)] \quad (\text{IV.46})$$

Suponiendo que los unos y los ceros son igualmente probables en un mensaje la ecuación (IV.46), nos da la probabilidad de error en la decodificación de cualquier dígito. Note usted que la probabilidad de error solamente depende de la relación A/σ , que es la relación entre la amplitud de los pulsos y la desviación estándar del ruido. La cantidad σ comúnmente se refiere como la amplitud *rms* del ruido. La relación A/σ se conoce como la relación señal a ruido *rms*.

En la figura IV.36 se grafica la probabilidad de error contra A/σ , en decibelios. A partir de esta gráfica notamos que para $A/\sigma = 7.4$ (17.4 dB) la P_e es 10^{-4} , lo que se interpreta como que en promedio 1 bit en 10^4 transmitidos puede detectarse erróneamente. Si 10^5 bits/s se transmiten, se deduce que habrá un error cada 0.1 s en promedio, lo que puede no ser satisfactorio. Sin embargo, si la señal se incrementa a $A/\sigma = 11.2$ (21 dB), es decir si se aumentan 3.9 dB, la P_e decrece a 10^{-8} . Entonces para 10^5 bits/s transmitidos habrá un error cada 1 000 s, o cada 15 minutos, lo cual es mucho más tolerable. En muchos casos de diseño práctico se utiliza una probabilidad de error de 10^{-5} , como un objetivo de diseño para sistemas de comunicación binarios.

Cabe aclarar que todo el análisis anterior lo hicimos suponiendo un código de línea unipolar sin retorno a cero y pulsos equiprobables; si estas condiciones cambian, por ejemplo cambiando el código de línea utilizado, también cambian en general los niveles de decisión y la probabilidad de error (por ejemplo como en el problema 2).

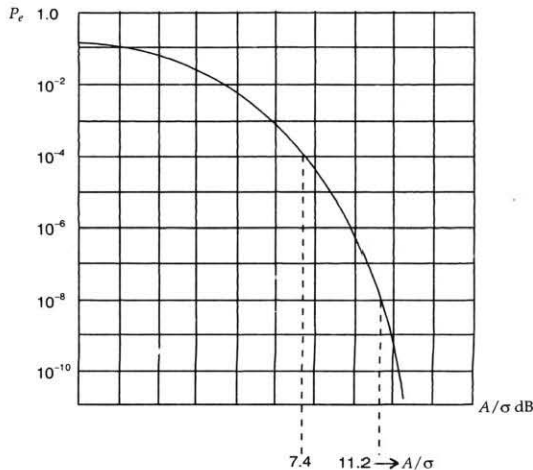


Figura IV.36. Probabilidad de error para detección binaria con ruido gaussiano aditivo

Arriba de $A/\sigma = 6$ o 16 dB (aproximadamente), la probabilidad de error decrece muy rápidamente con pequeños cambios de la señal. En el ejemplo analizado, incrementando la señal por un factor de 3.6 dB ($A/\sigma = 7.4$ a $A/\sigma = 11.2$) el error se reduce por un factor de 10^4 . La existencia de un pequeño intervalo de relaciones señal a ruido arriba del cual la razón de error es tolerable, pero que por arriba los errores ocurren con frecuencia, permite definir un efecto o nivel de cambio. La relación señal a ruido a la que este efecto toma lugar se le denomina el nivel de cambio (*threshold level*). Para la transmisión de dígitos binarios este nivel de cambio se selecciona en algún punto entre $A/\sigma = 6$ a $A/\sigma = 8$ (16 a 18 dB). Note que esto no implica supresión total de ruido arriba del nivel de cambio, solamente nos indica que para amplitudes de pulso mayores a 10 veces la amplitud *rms* del ruido, los errores en la transmisión de dígitos binarios ocurren a una tasa tolerable.

IV.7.1. Niveles de decisión óptimos

En el desarrollo de la probabilidad de error utilizamos varios argumentos más o menos intuitivos. Supusimos que los pulsos tienen la misma probabilidad de ocurrencia; arbitrariamente seleccionamos el nivel de decisión como $A/2$ para un código binario polar sin retorno a cero. ¿Podemos entonces preguntarnos cuál es el significado real de estas suposiciones?; ¿es posible reducir la probabilidad de error si se selecciona otro nivel de decisión?; ¿cómo se puede obtener una probabilidad de error mínima? En esta sección se analizarán estas preguntas y otras relacionadas con el mismo tema.

Para hacer esta explicación más cuantitativa, regresemos a la formulación original de la probabilidad de error. A partir de las ecuaciones (IV.39), (IV.41) y (IV.42) y sustituyendo en ellos un nivel de decisión arbitrario a , tenemos que

$$P_e = P_0 \int_a^{\infty} f_0(v) dv + P_1 \int_{-\infty}^a f_1(v) dv \quad (IV.47)$$

remarcamos que la ecuación previa es válida para el caso de señales unipolares sin retorno a cero. Una óptima selección del nivel de decisión a corresponde a minimizar la probabilidad de error P_e ; como P_e es una función de a , entonces simplemente diferenciando respecto a a encontramos el nivel de decisión óptimo. En particular para este caso, obtenemos que

$$\frac{\partial P_e}{\partial a} = 0 = -P_0 f_0(a) + P_1 f_1(a)$$

o

$$\frac{f_1(a)}{f_0(a)} = \frac{P_0}{P_1} \quad (IV.48)$$

Aparentemente el nivel de decisión óptimo a (en el sentido de mínima probabilidad de error) depende de la forma de las dos funciones de densidad de probabilidades $[f_0(v)$ y $f_1(v)]$, así como de las probabilidades *a priori* P_0 y P_1 . Si $P_0 = P_1 = 1/2$, el nivel de decisión óptimo está determinado por el punto donde las funciones de densidad se intersectan. Para señales unipolares sin retorno a cero este punto es precisamente $a = A/2$, que se obtiene sustituyendo las ecuaciones (IV.39) y (IV.41) evaluadas en a y sustituyéndolas en la ecuación (IV.48). Para el caso $P_0 \neq P_1$, el nivel de decisión se

desplaza como es de esperarse. Específicamente para el ejemplo de señales unipolares con ruido gaussiano aditivo, la solución de la ecuación (IV.48) es (su desarrollo se deja como ejercicio para el lector)

$$a_{\text{opt}} = \frac{\sigma^2}{A} \ln \frac{P_0}{P_1} + \frac{A}{2} \quad (\text{IV.49})$$

como esperábamos, a se incrementa positivamente si $P_0 > P_1$, y negativamente si $P_0 < P_1$. El desplazamiento real depende de la amplitud de la señal, de la varianza del ruido y de la relación P_0/P_1 .

En la práctica la selección del nivel de decisión no es muy crítica, en el caso de señales unipolares con ruido blanco aditivo por lo regular se establece a $A/2$; específicamente, si $A/\sigma = 8$ y $P_0/P_1 = 3$ entonces $a_{\text{opt}} = \frac{\sigma}{8} + \frac{A}{2}$. Así, el nivel óptimo se desplaza del valor $a = A/2$ por una fracción de la amplitud *rms* del ruido, lo que en general es un cambio insignificante. Aunque podríamos pensar que para conservar la misma relación señal a ruido se podría incrementar la amplitud de la señal con el mismo incremento, eso no es práctico; por ejemplo, con la relación señal a ruido inicialmente establecida a 18 dB, eso representaría un desplazamiento de aproximadamente 0.13 dB, lo que a su vez provocaría un cambio despreciable en la probabilidad de error (véase figura IV.36).

IV.8. TÉCNICAS DE DETECCIÓN Y CORRECCIÓN DE ERRORES

Redundancia

La redundancia implica transmitir cada carácter dos veces. Si el mismo carácter no se recibe dos veces en sucesión entonces ha ocurrido un error en la transmisión. El mismo concepto puede usarse para los mensajes. Si la misma secuencia de caracteres no es recibida dos veces en sucesión, exactamente en el mismo orden, entonces ha habido un error en la transmisión.

Codificación de cuenta exacta

Con codificación de cuenta exacta, el número de unos en cada carácter es el mismo. Un ejemplo de código de cuenta exacta es el código ARQ mostrado en el cuadro IV.9. En el código ARQ cada carácter tiene tres unos, y por lo tanto una simple cuenta del número de unos recibidos puede determinar si ha ocurrido un error.

Paridad

El método más común para detectar errores es a través de la paridad. Un simple bit llamado de paridad se suma a cada palabra de tal manera que el número total de unos en la palabra, incluido el bit de paridad, sea par para paridad par o impar para paridad impar.

Supongamos que se desea utilizar paridad impar y se utiliza el código ASCII. El código para la letra es P100 0001, donde P representa el bit de paridad, ya que se desea transmitir paridad impar; P debe tomar un valor tal que la suma total de unos sea impar. Por lo tanto, para este ejemplo P debe tomar el valor de 1 y 1100 0001 debe ser la palabra transmitida.

El bit de paridad se inserta en el transmisor y la paridad se verifica en el receptor. El receptor debe conocer el tipo de paridad que se está utilizando. Si se selecciona una paridad impar y una

palabra junto con su bit de paridad tiene un número de unos par, eso significa que por lo menos ha habido un error en alguno de los bits. Una limitación con paridad es que si ocurre un error en dos bits de una palabra, este error no podrá detectarse. Hay muchos circuitos que pueden utilizarse como generadores y verificadores de paridad, el menos complejo de éstos aparece en la figura IV.37.

CUADRO IV.9. Código de cuenta exacta ARQ

Código binario	Carácter	
bit: 1234567	Letra	Figura
0001110	Letter shift	
0100110	Figure Return	
0011010	A	—
0011001	B	?
1001100	C	:
0011100	D	(WRU)
0111000	E	3
0010011	F	%
1100001	G	@
1010010	H	£
1110000	I	8
0100011	J	(bell)
0001011	K	(
1100010	L)
1010001	M	.
1010100	N	,
1000110	O	9
1001010	P	0
0001101	Q	1
1100100	R	4
0101010	S	'
1000101	T	5
0110010	U	7
1001001	V	=
0100101	W	2
0010110	X	/
0010101	Y	6
0110001	Z	+
0000111		(blank)
1101000		(space)
1011000		(line feed)
1000011		(carriage return)

Si en la palabra hay n bits se necesitan $n-1$ compuertas XOR para que el circuito funcione como generador de paridad. No importa cuántos bits haya en la palabra, la salida del circuito siempre será 1 si hay un número impar de unos en la palabra; la salida siempre será 0 si en la palabra hay un número par de unos; por lo tanto, si se desea una paridad par, la salida de este circuito puede usarse directamente para el bit de paridad. Si se desea una paridad impar, la salida de este circuito debe pasarse a través de un inversor para generar el bit de paridad apropiado. Para que este circuito pueda utilizarse como un verificador de paridad, la palabra más el bit de paridad deben aplicarse a la entrada del circuito. Si no han ocurrido errores, la salida siempre será baja para paridad par y alta para paridad impar. Lo que se haga cuando un error se detecta depende del sistema. En algunos sistemas se inicia una petición automática de retransmisión cuando se detecta un error de paridad; algunas veces el mensaje puede comprenderse a pesar del error. Algunos sistemas no solicitan retransmisión, en vez de eso despliegan el símbolo (?) para el carácter que se ha recibido con error.

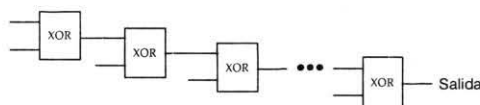


Figura IV.37. Circuito generador y verificador de paridad

Verificación de redundancia vertical y longitudinal

Si se utilizan dos bits de paridad, el error de un bit puede detectarse y corregirse. VRC (verificación de redundancia vertical) simplemente consiste en insertar y verificar el bit de paridad para cada palabra tal como se describió previamente. Un grupo de palabras se llama un bloque; los bits adicionales de paridad se calculan basándose en los primeros bits de cada palabra del bloque, en los segundos bits de cada palabra del bloque y así sucesivamente; estos bits forman la LRC (verificación de redundancia longitudinal).

Como ejemplo, supongamos que se utiliza paridad impar para la VRC y paridad par para la LRC, y consideremos el siguiente bloque de palabras que corresponden al mensaje *THE CAT*, donde cada letra se codifica en código ASCII, como se muestra en la figura IV.38.

El bit (*) se calcula como un bit VRC. Para nuestro ejemplo supongamos que el bit 2 de la letra *E* se recibió erróneamente como 0. El VRC puede detectar un error de paridad para este carácter, pero por sí mismo no puede especificar el bit que se recibió con error. El LRC detecta también un error en la palabra de los segundos bits, pero también por sí mismo no puede identificar en qué letra está el error. Sin embargo, en conjunto el LRC y el VRC (por medio de la intersección) identifican el bit que

	T	H	E	Space	C	A	T	L R C
LSB	0	0	1	0	1	1	0	1
	0	0	0	0	1	0	0	1
bit 2 →	1	0	1	0	0	0	1	1
	0	1	0	0	0	0	0	1
	1	0	0	0	0	0	1	0
	0	0	0	1	0	0	0	1
MSB	1	1	1	0	1	1	1	0
VRC	0	1	0	0	0	1	0	(*) = 0

Figura IV.38. Ejemplo de verificación de redundancia vertical y longitudinal

produjo el error: el bit dos de la letra *E*, por lo que una inversión de este bit corrige el error. Sin embargo, al igual que el método simple de paridad, un error en dos bits tanto en LRC como en VRC no es posible detectarlos.

Verificación de redundancia cíclica

La verificación de redundancia cíclica (CRC) se usa normalmente con transmisiones que utilizan el código de caracteres EBCDIC. El procedimiento CRC consiste en dividir el tren de datos entre un polinomio generador $G(x)$; el cociente se descarta, y el residuo se transmite como el CRC. El número de bits en el CRC es igual al mayor exponente de la función generadora, que está definida para EBCDIC en la recomendación V.41 de la UIT, a saber

$$G(x) = x^{16} + x^{12} + x^5 + 1 \quad \text{donde } x^0 = 1$$

esta función también se conoce como el CRC-16 (debido a su exponente de más alto orden) y tiene una longitud de 16 bits o dos bytes. Estos dos bytes del CRC se conocen como la secuencia de verificación de bloque (BCS) o la secuencia de verificación de trama (FCS). Los correspondientes caracteres EBCDIC se denominan bloque de caracteres verificados (BCC) o la trama de caracteres verificados (FCC).

Los exponentes de la función $G(x)$ identifican la posición de los bits que contienen un 1, por lo que $G(x)$ puede escribirse como

$$\begin{array}{ccccccc} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ b_{16} & & b_{12} & & & & & & b_5 & & & & & & & & b_0 \end{array}$$

El proceso de división mencionado anteriormente no se hace de la manera usual. En la parte de sustracción de la división, el residuo se obtiene efectuando una operación XOR, en vez de la sustracción. Por lo que $G(x)$ dividirá a otro número con un número igual de dígitos sin importar si $G(x)$ es mayor o menor; el criterio para decidir el valor de la división es: si $G(x)$ tiene igual número de dígitos que el dividendo, entonces el resultado es 1, en caso contrario es cero.

A modo de ejemplo supongamos el mensaje

$$\begin{aligned} M(x) &= x^7 + x^5 + x^4 + x^2 + x + 1 \\ &= 10110111 \end{aligned}$$

por simplicidad, usaremos $G(x) = x^5 + x^4 + x + 1$, $M(x)$ se multiplica primero por el número de bits del CRC (es el mismo que el mayor exponente de $G(x)$), con lo que se obtiene

$$x^5 (x^7 + x^5 + x^4 + x^2 + x + 1) = x^{12} + x^{10} + x^9 + x^7 + x^6 + x^5 = 1011011100000$$

que ahora dividimos entre $G(x)$, efectuando la operación como se explicó anteriormente; el CRC se agrega al mensaje produciendo la siguiente secuencia de bits

$$10110111 \quad 01001$$

En el receptor esta secuencia de bits se *divide* entre el mismo polinomio $G(x)$ utilizado

$$\begin{array}{r}
 \overline{1010111} \rightarrow \text{Se descarta} \\
 110011 \overline{1011011100000} \\
 \underline{110011} \\
 111101 \\
 \underline{110011} \\
 111010 \\
 \underline{110011} \\
 100100 \\
 \underline{110011} \\
 101110 \\
 \underline{110011} \\
 111010 \\
 \underline{110011} \\
 01001 = \text{CRC}
 \end{array}$$

en el transmisor y si el residuo es cero, no hubo error.

$$\begin{array}{r}
 \overline{1010111} \\
 110011 \overline{1011011101001} \\
 \underline{110011} \\
 111101 \\
 \underline{110011} \\
 111010 \\
 \underline{110011} \\
 100100 \\
 \underline{110011} \\
 101110 \\
 \underline{110011} \\
 111010 \\
 \underline{110011} \\
 00000 \quad \text{No hay error}
 \end{array}$$

Código Hamming

El código Hamming es un código de corrección hacia adelante, lo que significa que el receptor tiene la capacidad para detectar y corregir el error. El número de bits en el código Hamming depende del número de bits que contenga la cadena de bits que se van a transmitir. Si m representa el número de bits contenidos en la cadena, y n representa el número de bits en el código Hamming, entonces n es el entero más pequeño que cumpla con

$$2^n \geq m + n + 1 \quad (\text{IV.50})$$

los bits de Hamming pueden insertarse en cualquier posición de la cadena de bits. Aunque estas posiciones son arbitrarias, deben decidirse de antemano y permanecer fijas durante toda la transmisión, además de que deben ser conocidas por el transmisor y el receptor.

A manera de ejemplo supongamos que la cadena de bits de datos contiene 12 bits

1 0 1 1 0 0 0 1 0 0 1 0

como $m = 12$, se determina a partir de (IV.50) que n debe ser igual a 5 ($n = 5$). Los bits de Hamming (bits H) se colocan en las posiciones previamente seleccionadas (remarcamos que estas posiciones son arbitrarias).

17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
H	1	0	1	H	1	0	0	H	H	0	1	0	H	0	1	0

Para determinar el valor de los bits de Hamming se hace una operación XOR con todas las posiciones de bits de datos que contengan unos; las posiciones se escriben como un número binario de longitud igual a n ($n = 5$ para este caso).

2	00010
6	<u>00110</u>
	00100
12	<u>01100</u>
	01000
14	<u>01110</u>
	00110
16	<u>10000</u>
	10110 Bits de Hamming

Estos bits de Hamming se colocan en las posiciones de los bits H de izquierda a derecha, entonces la cadena de datos codificados que se transmite es:

1 1 0 1 0 1 0 0 1 1 0 1 0 0 0 1 0

Supongamos que esos datos se recibieron con error en el bit 14, como se muestra a continuación

1 1 0 "0" 0 1 0 0 1 1 0 1 0 0 0 1 0

para detectar el error se extraen los bits de Hamming y se hace una operación XOR con todas las posiciones de bits de datos que contengan unos.

	10110 bits de Hamming
2	<u>00010</u>
	10100
6	<u>00110</u>
	10010
12	<u>01100</u>
	11110
16	<u>10000</u>
	01110 = 14 posición del error

Si ocurre un error en los bits de Hamming, o si hay más de un error, este esquema no es capaz de detectarlo y corregirlo.

En resumen, cualquier esquema para detectar y corregir errores involucra la transmisión de bits adicionales. El objetivo es eliminar la necesidad de retransmisión, pero la inclusión de bits adicionales para detectar y corregir errores incrementa el tiempo de transmisión. De aquí que para la implementación de cualquier esquema de corrección de errores hacia adelante (FEC), se debe comparar la frecuencia de los errores y el tiempo de retransmisión con el incremento de tiempo de transmisión provocado por la inserción de bits adicionales para FEC.

IV.9. MODULACIÓN DELTA

Como ya mencionamos en las secciones anteriores, las técnicas de modulación PCM se usan mucho en la transmisión digital de voz (telefonía), lo mismo que para la transmisión de diversos tipos de señales de telemetría. Esta técnica también se ha utilizado para la transmisión de imágenes tanto fijas como de video, así como para diversas señales analógicas convertidas a un formato digital por medio de un proceso de conversión A/D. Es importante hacer notar que se ha fomentado el uso de transmisión y procesamiento digital de señales aunado al desarrollo de equipo de computación, a pesar de que se tiene que hacer una conversión A/D sobre la señal; no obstante, la transmisión digital de señales generalmente resulta en una expansión del ancho de banda: si el número de niveles de cuantificación o, equivalentemente, el número de bits por muestra se incrementa, el ancho de banda de transmisión también se incrementará de forma proporcional.

Se han propuesto métodos alternativos para convertir señales analógicas a un formato digital para algunas aplicaciones, con el fin de reducir el ancho de banda de transmisión requerido, mejorar la calidad del sistema y reducir costos. La modulación delta es uno de estos esquemas de modulación que se han adoptado para algunas aplicaciones como la transmisión de voz o de imágenes. En estas aplicaciones hay una gran cantidad de redundancia en la información de la señal que va a ser transmitida. La señal transmitida (pasada) puede proporcionar información para tratar de predecir el comportamiento futuro de la señal, de manera tal que las nuevas señales necesitan únicamente transmitirse si ocurre un cambio significativo en la señal.

Consideremos el diagrama a bloques de un sistema de modulación delta tal como se muestra en la figura IV.39. Igual que en la conversión A/D, la señal analógica $x(t)$ debe muestrearse previamente. Los valores de las muestras, designadas por x_j , $j = \dots, -2, -1, 0, 1, 2, 3 \dots$ se comparan con valores predichos de las muestras denotadas por g_j , y la diferencia e_j es la que se pasa al cuantificador. Obviamente, si e_j es pequeña para la mayoría del tiempo, tal que la predicción sea buena, entonces se necesitarán pocos bits para representar la señal de diferencia e_j . Un modulador delta usa un cuantificador de dos niveles, de tal manera que sólo se necesita un bit para representar la señal de diferencia. En la figura IV.38 los dos niveles del cuantificador se representan por $\pm k$. En el receptor la señal de diferencia cuantificada se suma a la salida del predictor para obtener una estimación discreta x_{je} de la muestra deseada x_j . Esta señal estimada discreta se pasa a través de un filtro pasa bajas para suavizarla y generar la señal estimada deseada $x_e(t)$.

Aparentemente para mejorar la calidad de este esquema deben usarse más de dos niveles en el cuantificador. Este esquema más general se conoce como *Modulación por pulsos codificados diferencial* (DPCM). Por lo tanto, modulación delta es un caso especial de DPCM que se obtiene cuando se utilizan sólo dos niveles en el cuantificador.

El predictor que se usa generalmente es una suma pesada de un cierto número de muestras estimadas anteriores (pasadas). En general la forma de un predictor lineal de k -ésimo orden

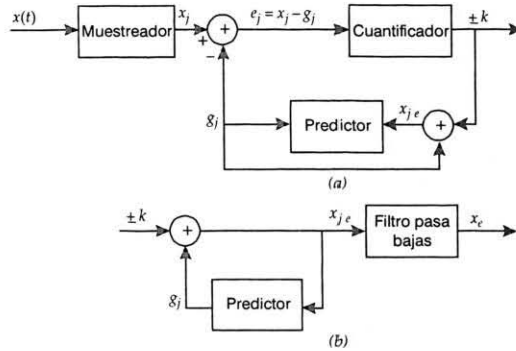


Figura IV.39. Modulación delta, a) transmisor, b) receptor

(utiliza para calcular la muestra estimada actual, k , muestras estimadas anteriores) puede escribirse como

$$g_j = \sum_{l=1}^k h_l x_{ej-l} \quad (\text{IV.51})$$

los coeficientes h_l son los factores de peso. El predictor más simple es el que usa para estimar la muestra actual (x_j) únicamente una muestra estimada anterior, para este caso $g_j = h_1 x_{ej-1}$.

Los coeficientes $h_l, l = 1, 2 \dots k$ pueden seleccionarse para reducir, lo más posible, los errores de las muestras estimadas comparadas con las muestras reales. Por ejemplo, una medida común de la calidad del sistema es el error cuadrático medio, promediado sobre muchas muestras o promediado estadísticamente sobre la estadística (si es conocida) de las muestras entrantes x_j .

El predictor lineal descrito por la ecuación (IV.51) es un ejemplo de un filtro digital no recursivo que puede efectuarse con elementos de retraso. La figura IV.40 ilustra un diagrama a bloques del predictor. Los bloques Z^{-1} representan elementos que retrasan las muestras entrantes por un intervalo de muestreo.

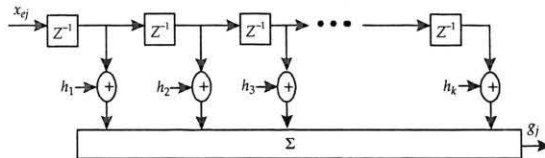


Figura IV.40. Predictor para modulación delta (filtro digital no recursivo)

Como en el caso de modulación PCM, en modulación delta también aparece una distorsión de cuantificación, sin embargo la forma para reducirlo es diferente; en el caso de PCM esta reducción se obtiene aumentando el número de niveles de cuantificación, en este caso la distorsión puede reducirse únicamente aumentando la frecuencia de muestreo. Se pueden encontrar sistemas de modulación delta que utilizan frecuencias de muestreo mucho mayores que las utilizadas en

modulación PCM. Esto produce que la razón de bits sea mayor a la que originalmente se esperaba tener con dos niveles de cuantificación. En algunos casos la razón de bits requerida puede ser mayor que la utilizada por un sistema PCM. Para solucionar esa dificultad se utilizan esquemas de modulación delta adaptativa, los cuales serán discutidos en la próxima sección.

En adición al ruido de cuantificación encontramos otro tipo de ruido en modulación delta: éste se produce cuando los niveles $\pm k$ son demasiado pequeños para seguir cambios bruscos de la señal; a este tipo de ruido se le conoce como error por sobrecarga. En la figura IV.41 se muestran los dos tipos de ruido que aparecen en modulación delta; es obvio que la condición para evitar el error por sobrecarga es

$$\left| \frac{dx(t)}{dt} \right|_{\max} \leq \frac{\Delta V}{T_s} \quad (\text{IV.52})$$

donde se tomó ΔV en general como el tamaño del intervalo de cuantificación.

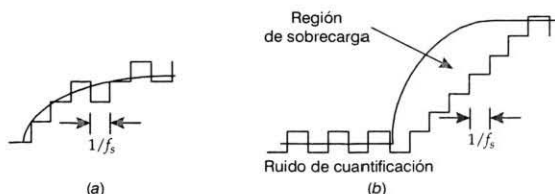


Figura IV.41. Fuentes de error en modulación delta. a) Ruido de cuantificación, b) error por sobrecarga

IV.9.1. Modulación delta adaptativa (ADPCM)

Como mencionamos, la calidad del esquema de modulación delta depende de nuestro conocimiento de la estadística de la señal. Si tenemos claramente definida la señal de entrada podremos diseñar predictores invariantes en el tiempo (es decir con coeficientes de peso fijos). En muchos casos la estadística de la señal no está bien conocida, y es necesario un esquema de codificación adaptativa (o sea con factores de peso que varíen con la estadística de la señal).

El término ADPCM, en general se utiliza para definir dos diferentes esquemas: adaptación de los niveles de cuantificación y adaptación del predictor. Los cuantificadores adaptativos ofrecen una mejora en la razón S_o/N_q de entre 3 y 7 dB sobre esquemas no adaptativos. La calidad de la adaptación depende de la calidad del predictor, que puede determinar la media de la señal o en algunas aplicaciones su varianza. La realización depende del compromiso entre la complejidad del sistema y su facilidad de realización.

Las aproximaciones más comunes para sistemas de codificación de voz combinan cuantificadores adaptativos con predictores adaptativos. Con sistemas de predicción adaptativa se puede incrementar la razón S_o/N_q en un factor de 10. La recomendación G.721 de la UIT para ADPCM emplea adaptación con retroalimentación tanto para el predictor como para el cuantificador; en la figura IV.42 se muestra el diagrama a bloques tanto del codificador ADPCM como del decodificador. La entrada al codificador es un canal E0 a 64 Kbits/s que transporta las muestras de voz codificadas en PCM normal ($f_s = 8$ kHz y 8 bits por muestra con compansión). El codificador transforma la señal PCM a 64 Kbits/s a una señal ADPCM a 32 Kbits/s, por lo que usualmente éste se conoce como transcoder.

El bloque de conversión de formato en el codificador reconstruye las palabras PCM comprimidas a un formato lineal, inversamente en el decodificador este bloque efectúa la compresión de la señal lineal ya sea a *ley A* o a *ley μ* .

Un cuantificador de 4 bits se utiliza para cuantificar la señal de diferencia. Cabe hacer notar que en el lazo de retroalimentación del codificador se tiene un cuantificador adaptativo inverso que reconstruye la señal de diferencia. El predictor, entonces, utiliza el valor de la muestra previa estimada $s_e(n)$ para calcular el próximo valor de la muestra estimada, el cual es retroalimentado y sustraído de la entrada para formar la señal de entrada diferencial.

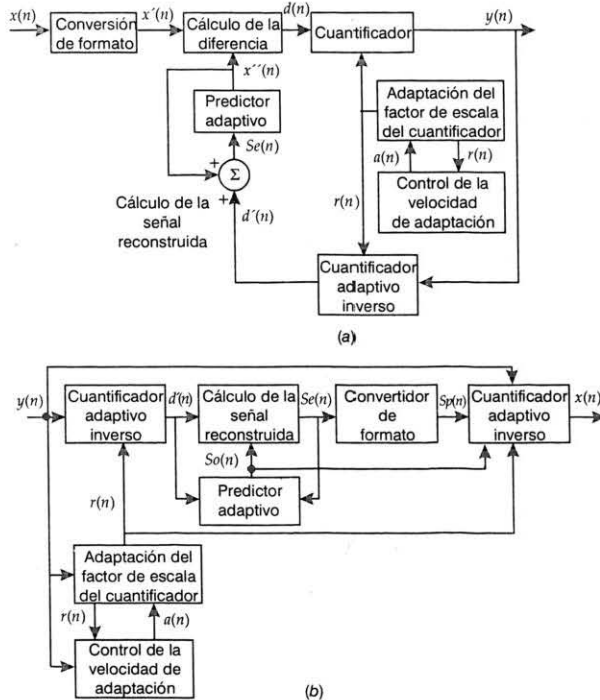


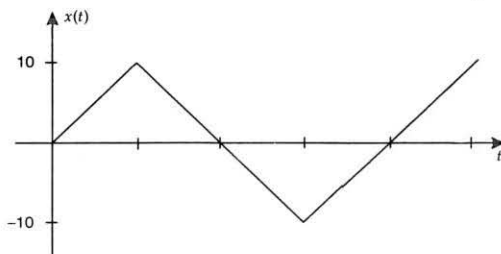
Figura IV.42. Modulación ADPCM recomendación G.721 de la UIT. a) Codificador, b) decodificador

IV.10. PROBLEMAS

- IV.1. Calcule los niveles de cuantización óptimos y dibuje la función de transferencia del cuantizador, si los límites de las integrales de la ecuación (IV.12) se cambian a $f_i - (d/2)$ y $f_i + (d/2)$.
- IV.2. En el código bipolar los unos se codifican como $+A/2$ y los ceros como $-A/2$.
- Determine la probabilidad de error.
 - Determine el nivel óptimo si $P_1 = P_0$.
 - Demuestre que el nivel óptimo para $P_1 \neq P_0$ es igual a:

$$a_{\text{opt}} = \frac{\sigma^2}{A} \ln \frac{P_0}{P_1}$$

- IV.3. El error de cuantificación de un sistema PCM de cuantificación uniforme no debe ser mayor que $\pm 5\%$ del valor pico a pico de $x(t)$. Encuentre el número mínimo de dígitos binarios que conforman las palabras de código.
- IV.4. Se desea transmitir mediante un sistema PCM binario, de cuantificación uniforme, una señal cuyo espectro de frecuencia no supere los 3 600 Hz; la tasa máxima de señalización se fijó en 40 Kbits/seg. Dibuje un diagrama a bloques del transmisor y especifique los parámetros del diseño.
- IV.5. Un convertidor analógico-digital se excita con la señal de la figura. Calcule:
- El rango dinámico requerido por el convertidor A/D.
 - La relación señal a ruido en función del número de bits en los que se codifica cada muestra.



- IV.6. Un convertidor analógico-digital tiene una señal de entrada dada por

$$x(t) = \sum_{m=-\infty}^{\infty} X_n e^{-j\omega_0 t}$$

Demuestre que

$$\frac{S}{N_Q} = \frac{12}{D^2} \cdot 2^{2n} \sum_{m=-\infty}^{\infty} |X_n|^2$$

donde D es el rango dinámico.

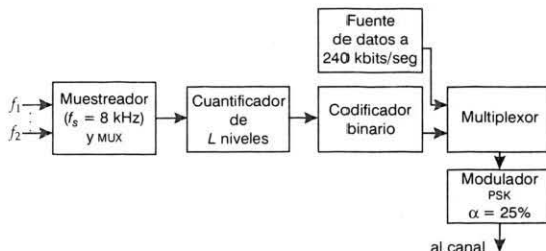
IV.7. Un canal de transmisión con un ancho de banda de 1 MHz está disponible para transmitir información.

- Un sistema PCM y una fuente de datos a 240 Kbits/seg se multiplexan en tiempo. La señal de salida se alimenta a un modulador PSK (se utiliza un factor de *roll-off* de 25%). Encuentre el número máximo de niveles de cuantización que pueden ser usados.
- El sistema PCM del inciso *a*) debe tener 256 niveles de cuantización y transmitir cuatro canales de 240 Kbits/seg cada uno; además se debe utilizar el mismo canal de transmisión. Indique qué cambios deben efectuarse en el modulador para satisfacer estos requisitos.

NOTA: La fórmula para calcular la velocidad de transmisión en sistemas de modulación digital (FSK, ASK, QAM, etc.) que utilizan filtros senoidales es la siguiente:

$$\frac{1}{T} = \frac{n \times B_{\text{canal}}}{(1 + \alpha)}$$

donde n = número de bits por símbolo.



IV.8. Calcule la probabilidad de error y el nivel de decisión óptimo para el código bipolar. En el código bipolar se emplea la siguiente regla de codificación:

Los 1 se transmiten como pulsos de amplitud $+V$
 Los 0 se transmiten como pulsos de amplitud $-V$

IV.9. La letra A en el código ASCII se representa por '100 0001':

- Determine cuántos bits Hamming se requieren para encontrar un error en la transmisión.
- Inserte el número de bits Hamming necesarios colocando todos los bits Hamming a la derecha del código. A partir de esto determine cuál es el valor correcto que toman los bits Hamming.

b_n ----- b_5 ----- b_0
 1 0 0 0 0 1 H----H

- Supóngase que el bit cinco (b_5) se recibió con error. Muestre cómo se identifica la posición del error utilizando el código Hamming.

IV.10. En un sistema de modulación delta se tiene que $\Delta V = 0.1$ volt y $T_s = 20$ microsegundos. Con base en la examinación del mensaje

$$f(t) = A \left[1.2 \cos(400 \pi t + \theta_1) + 1.3 \cos(600 \pi t + \theta_2) + 0.5 \cos(1200 \pi t + \theta_3) \right]$$

donde θ_1 , θ_2 y θ_3 son ángulos de fase que pueden tener cualquier valor. Encuentre el mayor valor de A que no produce ruido de sobrecarga.

- IV.11. Una señal comprendida en el rango de frecuencias de 100 a 4 000 Hz con un $V_{pp} = 3$ volts se muestrea a una frecuencia de 8 kHz. Las muestras son cuantizadas en 64 niveles uniformemente espaciados. Calcule el ancho de banda de la señal PCM y la razón S/N en dB (suponga señales senoidales) si las muestras cuantizadas son transmitidas como dígitos binarios o como pulsos de 4 niveles.
- IV.12. Dibuje la forma de onda del siguiente patrón de datos:

0 1 1 0 0 0 1 0 0 0 0 0 0 0 1 1 0

- a) Si se transmite en código bipolar (AMI).
 b) Si se transmite en código B3ZS (sustitución binaria de tres ceros); analice los dos casos de paridad inicial.

NOTA: Codifique el primer 1 como un pulso +.

Polaridad del pulso precedente	Número de pulsos bipolares (unos) después de la última sustitución	
	Par	Impar
-	00-	+0+
+	00+	-0-

- IV.13. En un canal de transmisión encontramos que la amplitud *rms* del ruido es igual a 500 mV. ¿Cuál debe ser la amplitud de los pulsos para que el error en el receptor sea aproximadamente de 1×10^{-6} ?

NOTA: Utilice el siguiente cuadro para calcular $\text{Erfc}(x)$.

x	$\text{Erfc}(x)$
0.0	1.0
1.2	0.8969×10^{-1}
1.9	0.7210×10^{-2}
2.4	0.6885×10^{-3}
2.8	0.7501×10^{-4}
3.5	0.7431×10^{-6}
4.0	0.1542×10^{-7}

- IV.14. El ancho de banda de una señal de TV es de 4.2 MHz. Si el número de niveles de cuantización es de 512, determine:

- a) La velocidad de la señal PCM en bits/seg.
 b) Su ancho de banda.
- IV.15. Una señal $m(t)$ se transmite mediante PCM binario sin compresión. Si la relación S/N_q requiere que sea cuando menos de 50 dB, determine el número mínimo de niveles de cuantización requerido (suponga que $m(t)$ es senoidal).
- IV.16. Dibuje la forma de onda para el siguiente patrón de datos:

1 1 0 0 0 1 1 0 1 0 0 1 0 1 0

en código duobinario. En este código un 0 se codifica como cero volts y un 1 como un pulso $+V$ o $-V$ de acuerdo con la siguiente regla:

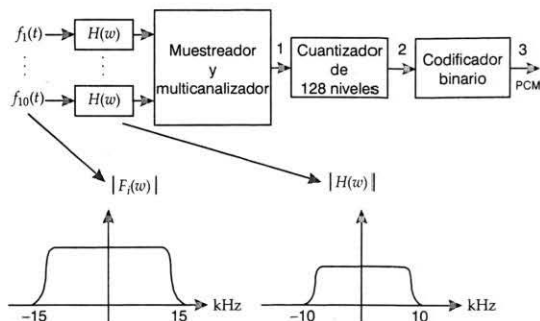
- Un 1 se codifica como un pulso de la misma polaridad que se usó para codificar el 1 que le preceda si los dos 1 se encuentran separados por un número par de 0.
- De otra forma se codifica mediante el negativo del pulso que se usó para codificar el 1 anterior.

NOTA: Considere el número cero como par y codifique el primer 1 como pulso $+V$.

- IV.17. Mediciones de temperatura que cubren el rango de -40°C a $+40^\circ\text{C}$, con $1/2^\circ\text{C}$ de exactitud, son tomadas a intervalos de 1 seg, las cuales son codificadas en formato binario para transmisión por PCM. ¿Cuál es la velocidad en bits/seg de la señal PCM?

NOTA: $\log_a x = \ln x / \ln a = \log x / \log a$

- IV.18. Un canal de información transporta señales de voz en el rango de 50 a 3 300 Hz. La señal se muestrea a una frecuencia de 8 kHz, los pulsos que resultan se transmiten sobre un sistema PCM.
- a) En el sistema PCM las muestras son cuantizadas en ocho niveles y se transmiten como dígitos binarios. Encuentre el ancho de banda de transmisión de la señal PCM.
- b) Repita el inciso a) si se utilizan 128 niveles de cuantificación.
- c) Calcule la razón señal a ruido de cuantificación en ambos casos (suponiendo señales senoidales).
- IV.19. Una señal se muestrea, cuantiza y codifica en PCM. Cada muestra se codifica en una *palabra* que consta de tres pulsos de información (datos) más uno de sincronía. Los pulsos de información pueden tener 4 niveles posibles. La transmisión se lleva a cabo sobre un canal cuyo ancho de banda es de 6 000 Hz; en el receptor se utiliza un filtro senoidal con un *roll-off* de 50%. Encuentre:
- a) La máxima razón de pulsos de la señal PCM.
- b) La correspondiente razón de información (en bits/seg) de la señal PCM.
- c) El ancho de banda máximo de la señal analógica.
- IV.20. Considere el sistema PCM que aparece en la siguiente figura. ¿Cuál es la mínima razón de muestreo? (suponga que primero se muestrea y luego se multiplexa). Ahora suponga que se toman 30 000 muestras por segundo; calcule la velocidad de transmisión (en bits/seg) de la señal PCM y el ancho de banda mínimo que se requiere en los puntos 1, 2 y 3.



- IV.21. Las lecturas de 100 sensores de presión, cuyo rango es de 10 a 50 *psi* y una resolución de 0.25 *psi*, se toman cada 10 seg. Las lecturas se transmiten usando técnicas de multicanalización PCM.
- ¿Cuál es la velocidad en bits/seg de la señal PCM?
 - ¿Qué ancho de banda se requiere para transmitir la señal PCM?
- IV.22. Se muestrean y multiplexan 10 señales, cada una con un ancho de banda de 10 kHz, a una razón de 25 000 muestras por seg. Cada muestra se codifica con 6 bits;
- Encuentre la velocidad de la señal PCM de salida en bits/seg. También calcule el ancho de banda necesario para transmitir la señal PCM.
 - Usando el mismo número de niveles de cuantización, cada muestra se transmite como una secuencia de pulsos de 4 niveles, ¿cuál es la velocidad de la señal PCM en pulsos por segundo y en bits por segundo? Calcule el ancho de banda de transmisión requerido.
 - Repita b) para el caso en que cada muestra cuantizada se transmite como un pulso multinivel sin mayor codificación.
- IV.23. Un voltaje $v(t) = 10 \sin [6.28t]$ se muestrea cada 0.33 seg, es decir, a una razón de 3 Hz. Las muestras son cuantizadas en 8 niveles uniformemente espaciados y codificadas en binario. Dibuje la señal original y abajo, a escala, las salidas del muestreador y el codificador. Calcule la razón señal a ruido de cuantización y los valores de los niveles óptimos de cuantización.

BIBLIOGRAFÍA

- [1] Winch, R. G., *Telecommunications Transmission Systems*, McGraw-Hill.
- [2] Miller, J. H. y V. S. Ahamed, *Digital Transmission Systems and Networks*, Computer Science Press.
- [3] Marvin, C. y E.I. Ewers, *Digital Signal Processing*, Texas Instruments.
- [4] Schwartz, Misha, *Information, Transmission, Modulation and Noise*, McGraw-Hill.
- [5] Peyton, Z. P., *Digital Communications Systems*, Prentice Hall.
- [6] Tomasi, W., *Sistemas de comunicaciones electrónicas*, Prentice Hall.

CAPÍTULO V

MODULACIÓN DIGITAL

V.1. INTRODUCCIÓN

LAS SEÑALES DE DATOS presentan en general la misma apariencia independientemente de la fuente que les dio origen. Es así cómo el flujo de dígitos que suministra una terminal de computadora no presenta diferencias apreciables, salvando la velocidad de señalización, de la que presenta una señal PCM o una imagen de TV digitalizada.

Estas señales digitales, tal como se presentan, son aptas para su transmisión sobre líneas o cables, es decir en banda base. Sin embargo, cuando se desea enviarlas a través de otros medios de transmisión es necesaria su adaptación al mismo de igual forma a la realizada en el caso analógico. Como sabemos, este proceso de adaptación recibe el nombre de modulación: en este caso, al ser la señal modulante una señal digital, surge el nombre, un tanto confuso, de modulación digital. Decimos confuso pues a los procesos de digitalización de señales analógicas tal como el PCM, DPCM, también se los denomina modulación; similar denominación suele aplicarse a las técnicas en las cuales se inscribe información analógica sobre un tren de pulsos, ya sea variando su amplitud (PAM), su posición (PPM) o su duración (PDM).

Con todo, estas señales donde la información está contenida en alguna característica de un tren de pulsos son sólo aptas para transmitirse en banda base. Consecuentemente, para adaptar una señal digital a un medio diferente de una línea deberá realizarse la modulación de una onda senoidal, también llamada modulación de onda continua (CW).

En resumen, modulación digital es el conocido proceso de modulación de una onda senoidal, con la salvedad de que la señal modulante es digital.

A lo largo de este capítulo consideraremos primero el concepto de eficiencia en la utilización del espectro aplicando el concepto de densidad de información. A continuación analizaremos los distintos tipos de modulación binaria y multisímbolo, considerando las ecuaciones fundamentales, la descripción en el espacio de señal, las realizaciones básicas de los moduladores y demoduladores y el espectro de frecuencia requerido para caso en particular. Finalmente se evaluará el comportamiento de cada esquema de modulación en presencia de ruido.

V.2. EFICIENCIA

Muchos sistemas de comunicaciones requieren antes de la transmisión procesamiento adicionales de la señal digital. Este proceso, indicado como adaptación, involucra primordialmente la traslación de la señal digital a otra región conveniente del espectro por medio de la modulación de una portadora. Este proceso de modulación presenta todas las posibles ventajas de la modulación analógica: facilidad de radiación, asignación de frecuencia, multiplexación, superación de limitaciones, etcétera.

Debe tenerse en cuenta que en la transmisión en banda base el ancho de la banda no es normalmente una limitación muy rígida puesto que la respuesta de los vínculos se deteriora en general gradualmente y siempre es posible ecualizarlos. En cambio, en los sistemas de radio enlace o en los canales telefónicos los anchos de banda se establecen rígidamente por criterios de asignación; en consecuencia, para lograr una mayor eficiencia en el uso de estos canales se requieren técnicas de señalización multinivel que en general están asociadas con el proceso de modulación.

En el caso concreto de los radios digitales se trata de obtener eficiencias similares a los radios analógicos. Así, por ejemplo, en las seis bandas de 2, 4, 6, 7, 8 y 11 GHz establecidas por el CCIR, para el caso analógico, se proveen 3.53 GHz de ancho de banda, donde se ubican unos 187 000 canales telefónicos, lo que significa poco menos de 19 KHz por cada canal.

Teniendo en cuenta que los canales telefónicos digitales requieren una tasa de señalización de 64 Kbits/s, resulta clara la necesidad de encontrar métodos convenientes que permitan aprovechar eficientemente el espectro y, en consecuencia, competir con los sistemas analógicos. Como luego veremos, estos métodos son las técnicas de modulación multisímbolo.

Un parámetro muy útil para caracterizar la eficiencia en la utilización de la banda en un sistema de modulación digital es la densidad de información

$$\delta = r/B \quad (V.1)$$

donde

r : tasa de señalización en bits/s

B : ancho de banda en Hz.

en consecuencia δ se expresa en bits/seg/Hz o simplemente en bits/ciclo, aunque suele decirse, impropiamente, bits/Hz.

Recordemos que teóricamente la máxima densidad de información para una señal binaria es 2 bits/seg/Hz, la cual se reduce a la unidad para modulación con doble banda lateral. Estos valores se incrementan mediante las técnicas multinivel, lográndose por ejemplo para el caso de líneas telefónicas no conmutadas, densidades de 3.1 bits/seg/Hz ($r = 9\,600$ bits/s), llegando en recientes realizaciones a 33 600 bits/seg.

V.3. MODULACIÓN BINARIA

Es oportuno comenzar el estudio de los sistemas de modulación digital para el caso binario y dejar la modulación multinivel para un apartado posterior.

En la modulación digital binaria de una onda continua encontramos las mismas tres variantes que en el caso analógico —amplitud, frecuencia y fase—, con la peculiaridad de que ahora conmutaremos alguno de esos parámetros entre dos valores posibles. Como luego veremos, hay casos particulares y ciertas variaciones de estas tres técnicas básicas.

V.3.1. Modulación en amplitud ASK

La modulación ASK probablemente es la primera técnica de modulación digital que se haya implementado prácticamente, puesto que se ha utilizado ampliamente (y aún hoy se sigue usando) para transmitir vía radio de telegrafía en código Morse.

En esta forma de modulación la amplitud de la portadora se varía entre dos niveles predeterminados en correspondencia con la señal binaria de datos; por lo regular uno de los dos niveles coincide con cero (condición OFF), por lo que en este caso suele denominarse la modulación ON-OFF (OOK).

Si denominamos $x(t)$ a la señal de datos, es decir a una secuencia de unos y ceros, la señal ASK será:

$$x_c(t) = A x(t) \cos(2\pi f_c t) \quad (V.2)$$

donde

$x(t) = 1$ o 0 sobre intervalos de T segundos.

f_c = frecuencia portadora.

En la figura V.1 se indica el diagrama a bloques de un modulador OOK, donde la función del filtro es reducir el nivel de los componentes indeseables y conformar la señal de salida.

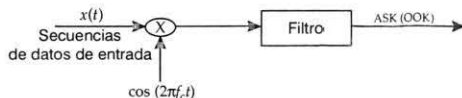


Figura V.1. Modulación ASK (OOK)

En la figura V.2 se muestran las formas de onda correspondientes. El efecto del filtro de salida puede tenerse en cuenta ya sea modificando la señal modulada $x_c(t)$ o por modificación de la señal modulante $x(t)$. En efecto, transformando la (V.2) tenemos:

$$X_c(f) = \frac{A}{2} [X(f - f_c) + X(f + f_c)] \quad (V.3)$$

o sea el espectro original desplazado a la frecuencia de portadora (fig. V.3). Esto significa que prácticamente el sistema ASK puede analizarse como un sistema de banda base y consecuentemente aplicar los resultados obtenidos en el capítulo V.

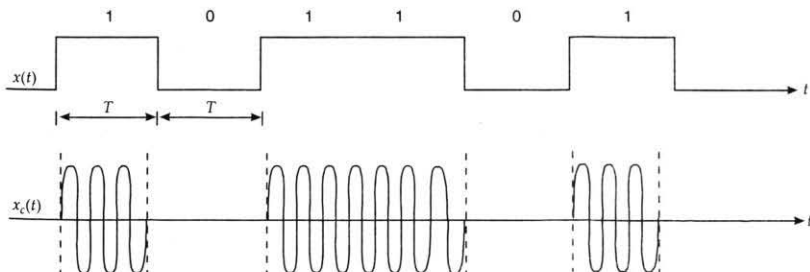


Figura V.2. Señales en modulación ASK (OOK)

Así, por ejemplo, si el procesamiento de la señal lo hiciéramos con un filtro de caída (*roll-off*) de seno elevado, el ancho de banda previo a la modulación sería

$$B = \frac{r}{2} (1 + \alpha) \text{ (Hz)} \quad (\text{V.4})$$

pasando a tener luego la modulación un ancho de banda de transmisión

$$B_T = 2 B = r (1 + \alpha) \text{ (Hz)} \quad (\text{V.5})$$

donde α es el factor de *roll-off* ya definido y r es la tasa de señalización.

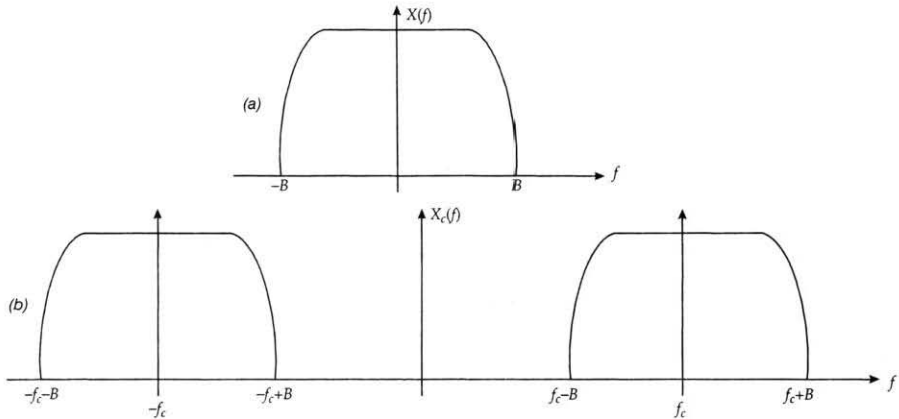


Figura V.3. a) Espectro de la señal modulante, b) espectro de la señal modulada en amplitud ASK

En lo referente a la detección de las señales ASK, en la figura V.4 aparece el diagrama a bloques de un receptor ASK.

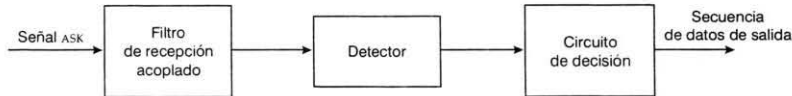


Figura V.4. Receptor ASK

El filtro de recepción normalmente es un filtro acoplado a los pulsos de RF que junto con el filtro del transmisor determinan la característica total requerida para todo el sistema. La detección puede hacerse de forma coherente (sincrónica) o no coherente; por lo regular se utiliza más esta última. Evidentemente la detección no coherente es de implementación más sencilla, pero, como veremos al considerar los efectos de ruido, provee una menor inmunidad cuando la señal de predetección presenta bajas relaciones señal a ruido.

La salida del detector, una vez filtrada para remover componentes indeseables, equivale a una

señal recibida en un sistema de banda base por la que le aplican todos los procesamiento de dichos sistemas: decodificación, muestreo y decisión.

V.3.2. Modulación en frecuencia FSK

La modulación por desplazamiento de frecuencia, FSK, provee una señal de amplitud constante que permite el uso de dispositivos saturables, lo que constituye una de sus ventajas respecto a ASK. La modulación FSK se usa principalmente en radios digitales de banda angosta, aunque también ha sido utilizada satisfactoriamente en equipos de radio digitales de banda ancha, principalmente aquellos que se han modificado a partir de radios analógicos que servían de soporte a sistemas multiplex FDM. Aparte de la ventaja mencionada, cabe apuntar su simplicidad, su bajo costo y, como veremos al final del capítulo, su comportamiento más estable que ASK ante desvanecimientos (*fading*).

Con todo, su comportamiento no resulta tan eficiente como PSK, aunque a diferencia de ésta permite una detección no coherente.

Es oportuno puntualizar que básicamente existen dos métodos de modulación digital de frecuencia: el primero se designa FSK, donde la señal digital $x(t)$ controla una llave que selecciona la frecuencia modulada de un banco de osciladores (dos en el caso binario); la señal modulada presenta discontinuidades en cada instante de conmutación a menos que la amplitud, la frecuencia y la fase de cada oscilador hayan sido cuidadosamente ajustadas. Es evidente que la dispersión del espectro de la señal modulada dependerá de dichas discontinuidades.

El segundo método de modulación de frecuencia intenta evitar esas discontinuidades y consecuentemente producir espectros más compactos; por eso se denomina CPFSK (*continuous-phase FSK*) y se basa en la modulación en frecuencia de un único oscilador por medio de la señal digital $x(t)$.

Consideremos primero el caso de FSK tradicional. La expresión de una señal binaria FSK es

$$x_c(t) = A \cos \left[2\pi (f_c + x(t) \cdot \Delta f) t \right] \quad (\text{V.6})$$

donde f_c es la frecuencia central (portadora virtual), $x(t)$ es la señal digital de banda base, simétrica NRZ de dos niveles, y Δf es la denominada desviación de frecuencia.

En la figura V.5 observamos una señal FSK típica, así como la correspondiente señal modulante.

Generalmente $f_c \gg 1/T$ (T duración del pulso), aunque en algunos sistemas, particularmente cuando el vínculo es una línea telefónica, son magnitudes del mismo orden. Así, por ejemplo, es

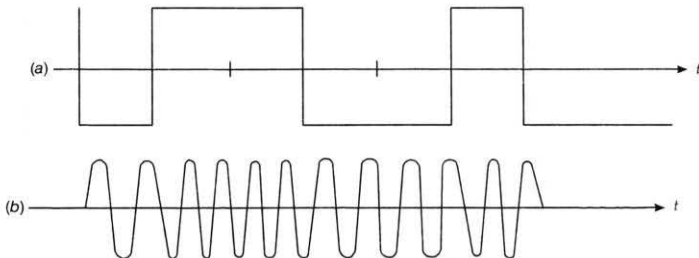


Figura V.5. Modulación FSK. a) Señal binaria de banda base, b) señal modulada

común para tasas de señalización de 1 200 bits/s que se utilice modulación fsk con portadora virtual de 1 700 Hz y desviación de 500 Hz.

El espectro de frecuencias de una señal fsk es, en general, difícil de obtener, por lo que primero consideramos un caso especial que permitirá obtener un panorama orientador [1]. Se supone que el mensaje binario es una secuencia alternada de unos y ceros, y que las dos frecuencias correspondientes son $f_1 = M/T$ y $f_2 = N/T$ (con M y N números enteros), ambas sincronizadas en fase. El espectro de la señal fsk que resulte puede visualizarse como la superposición de dos señales ook desplazadas relativamente en T , con lo que se obtiene que el espectro de frecuencias es:

$$X(f) = \text{sinc}[(f-f_1)T/2] + \text{sinc}[(f+f_1)T/2] + \text{sinc}[(f-f_2)T/2] + \text{sinc}[(f+f_2)T/2] \quad (\text{V.7})$$

donde $f_1 = f_c - \Delta f$ y $f_2 = f_c + \Delta f$.

En la figura V.6 representamos el lado positivo de la ecuación (V.7), donde se ha supuesto $\Delta f > 1/T$. El ancho de banda de esta señal resulta $2\Delta f + 2B$, donde $B = 1/(2T)$ es el ancho de banda de la señal de banda base.

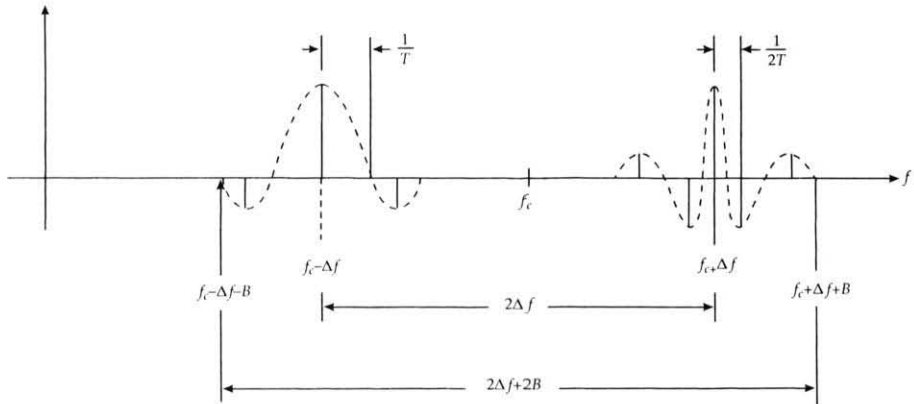


Figura V.6. Espectro de una señal modulada en fsk (parte positiva únicamente)

Dos casos extremos son de interés:

1) Si $\Delta f \gg B$, el ancho de banda se aproxima a $2\Delta f$. Esto corresponde si se utilizan tonos ampliamente separados en el sistema fsk; el ancho de banda es en esencia la separación de estos tonos, virtualmente independiente del ancho de banda de la señal binaria.

2) Si $\Delta f < B$, el ancho de banda se aproxima a $2B$. En este caso, aun con los tonos muy cercanos entre sí, el ancho de banda mínimo es el requerido para transmitir una señal ook; en este caso el ancho de banda está determinado por la señal digital.

En particular si la señal modulante es una secuencia arbitraria de pulsos binarios, cada uno conformado senoidalmente con un factor de *roll-off* (α), el ancho de banda aproximado de la señal fsk resulta

$$B_T = 2 \Delta f + r (1 + \alpha) \text{ (Hz)} \quad (\text{V.8})$$

En la generación de señales FSK binarias se incluye entonces un filtro de premodulación que permita lograr una mayor eficiencia al atenuar los lóbulos laterales de la señal modulada.

Precisamente en esta posibilidad para lograr un uso eficiente del espectro mediante un filtro de no muy difícil realización (premodulación) es donde reside la mayor competitividad de FSK.

El modulador puede ser lineal, de tipo VCO (oscilador controlado por voltaje), es decir un oscilador cuya frecuencia de salida es función de la tensión de entrada.

También hay moduladores del tipo digital que sintetizan las frecuencias requeridas; igualmente es posible utilizar dos osciladores sintonizados a $f_c \pm \Delta f$, cuyas salidas se conmutan según la señal de entrada. Este método cuya aplicación fue muy difundida en equipos de telegrafía armónica, tiene el inconveniente de producir transiciones de fase indeseables.

Los moduladores digitales se basan en circuitos divisores que cambian el factor de división en función de la señal de datos. La señal que se va a dividir está provista por un cristal que asegura la estabilidad. El diagrama a bloques de este dispositivo se indica en la figura V.7.

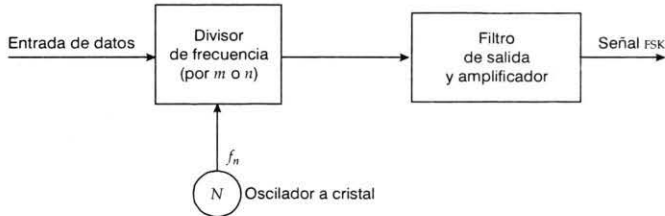


Figura V.7. Modulador FSK

Un amplificador y un filtro de salida completan el circuito de generación FSK. Las exigencias del circuito no son grandes pues su función sólo consiste en atenuar componentes indeseables ya que el espectro de salida ha sido conformado en banda base.

Las señales FSK pueden ser demoduladas coherente o incoherentemente. El detector coherente o de correlación requiere el conocimiento exacto de la fase de la señal de entrada, de ahí su nombre de coherente; su esquema se indica en la figura V.8.

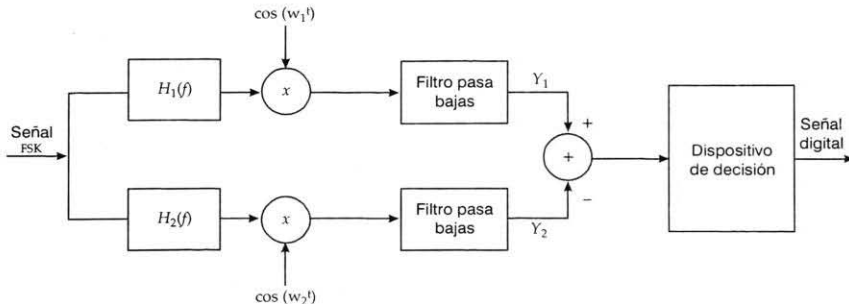


Figura V.8. Detector síncrono (o de correlación) para señales FSK

Cuando no conocemos exactamente la fase de la señal de entrada debemos recurrir a detectores de envolvente. En la figura V.9 se ejemplifica un detector de este tipo que emplea detectores de envolvente.

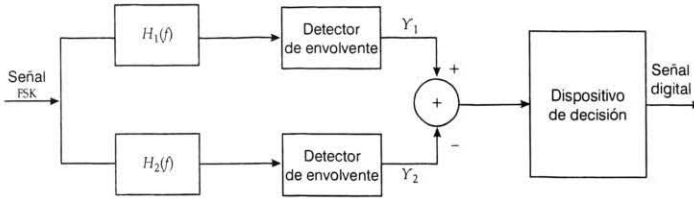


Figura V.9. Detector de envolvente para señales FSK

Supongamos que la entrada al demodulador es una señal $x_c(t)$, del tipo FSK de amplitud constante que instantáneamente conmuta entre dos frecuencias f_1 y f_2 . Si recordamos que esa señal puede considerarse como dos señales OOK entrelazadas, éstas se verán a la salida de cada filtro $H_1(f)$ y $H_2(f)$; es decir que $x_1(t)$ y $x_2(t)$ serán dos señales OOK tanto más nítidas cuanto $f_1 - f_2 > > 2T$, las cuales serán detectadas de acuerdo con el circuito de la figura V.4. La diferencia de ambos detectores proporcionará una señal bipolar que se analizará en el circuito de decisión.

V.3.3. Un caso especial MSK (minimum shift keying)

Partiendo de la ecuación (V.6) que define la modulación FSK, puede expresarse el par de señales $s_1(t)$ y $s_0(t)$ que representan los dígitos binarios en esta forma de modulación:

$$s_1(t) = A \cos(\omega_1 t + \theta) \quad (\text{V.9})$$

$$s_0(t) = A \cos(\omega_0 t + \theta)$$

donde $\omega_1 = 2\pi(f_c + \Delta f)$ y $\omega_0 = 2\pi(f_c - \Delta f)$. Obsérvese que ambos pulsos contienen igual energía, es decir

$$E = \int_0^T |s_i(t)|^2 dt = P_m T = \frac{A^2 T}{2} \quad (\text{V.10})$$

donde

P_m : potencia media.

T : duración de un bit.

E : energía de $s_0(t)$ o $s_1(t)$ durante T segundos.

En consecuencia, la ecuación (V.9) puede escribirse como

$$s_1(t) = \left(\frac{2E}{T}\right)^{1/2} \cos(w_1 t + \theta) \quad (V.11)$$

$$s_0(t) = \left(\frac{2E}{T}\right)^{1/2} \cos(w_0 t + \theta)$$

Es claro que al receptor le es más fácil reconocer cuál de los dos dígitos se le han enviado cuando éstos son más disímiles. El grado de semejanza entre dos señales se mide por el coeficiente de correlación γ que puede definirse como

$$\gamma = \frac{\int_0^T s_1(t) \cdot s_0(t) dt}{E} \quad (V.12)$$

Cuando $s_1(t)$ y $s_0(t)$ son muy semejantes, el coeficiente γ tiende a la unidad y ciertamente es la peor condición de trabajo para el receptor. Por otro lado, las señales donde

$$\int_0^T s_1(t) \cdot s_0(t) dt = 0$$

se denominan *ortogonales* y con ellas el receptor trabaja con mayor eficiencia, disminuyendo la probabilidad de errores. Sin embargo, la situación óptima que arroja la probabilidad de error menor se obtiene para $\gamma = -1$, que es la denominada condición antipodal, y sucede cuando $s_0(t) = -s_1(t)$. Aplicando ahora la expresión V.12 al par de ecuaciones (V.11) se tiene que

$$\gamma = \frac{\sin(w_1 - w_0) T}{(w_1 - w_0)} = \text{sinc}(4 \cdot \Delta f \cdot T) \quad (V.13)$$

que se representa en la figura V.10.

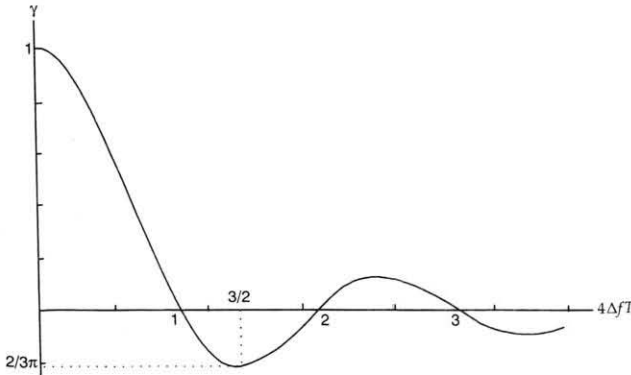


Figura V.10. Valores de γ en función de la desviación de frecuencia en modulación fsk

Obsérvese que el caso antipodal no existe en este sistema de modulación y que se acerca a la situación ortogonal a medida que crece la separación entre frecuencias; no obstante podemos ver que el valor más negativo del coeficiente de correlación sucede para $4 \Delta f T \approx 3/2$ o sea para una separación $2 \Delta f = 3/4 T$. En este caso podemos obtener igual probabilidad de error que en la condición ortogonal con señales 1.5 dB más bajas e igual situación de ruido.

Es decir, que colocando las señales en la separación establecida obtenemos una ligera ventaja en comparación con el caso de amplia separación espectral, para lo cual debe utilizarse en el receptor un par de detectores coherentes, sincronizados con el ángulo θ de la ecuación (V.9). Por este motivo a este esquema de señalización se le da el nombre de FSK coherente. Nuestro caso de interés se produce, sin embargo, para el primer cruce por cero. Cuando $4 \Delta f T = 1$, es decir para una separación mínima dada por

$$2 \Delta f = f_1 - f_0 = \frac{1}{2 T} \quad (\text{V.14})$$

sea la mitad de la tasa de señalización. Esto significa que en un intervalo T un cero produciría f_0 periodos de la portadora y un uno produciría $f_0 + 0.5 T$, o sea una diferencia de 180° . Más aún, la señal MSK mantiene continuidad de fase en las transiciones, por lo que pertenece a una clase de señales FSK denominadas de fase continua (CPFSK). Nótese que para cada uno de la portadora acumula 180° más que para cada cero (720° y 540° respectivamente en este ejemplo). Como consecuencia, MSK puede considerarse como un caso especial de FSK aunque su espectro corresponde a FSK.

La ecuación (V.14) nos muestra además que existe una relación coherente entre las frecuencias transmitidas y la tasa de señalización requerida.

Ahora demostraremos que la señal MSK puede generarse utilizando conformadores de onda senoidales. A partir de la ecuación (V.14) tenemos

$$f_0 = f_c - \frac{1}{4 T} = f_c - \Delta f$$

$$f_1 = f_c + \frac{1}{4 T} = f_c + \Delta f$$

con lo que describimos la señal MSK como

$$S_{\text{MSK}}(t) = A \cos [2 \cdot \pi (f_c \pm \Delta f) t]$$

quedando finalmente

$$S_{\text{MSK}} = A \cos (\pm 2 \pi \Delta f t) \cos (2 \pi f_c t) - A \sin (\pm 2 \pi \Delta f t) \sin (2 \pi f_c t) \quad (\text{V.15})$$

cuya realización se indica en la figura V.11a, donde la entrada de datos luego de atravesar el convertidor serie a paralelo se introduce en los conformadores senoidales. En el canal denominado I el conformador genera $\cos (\pm \pi t/2 T)$, mientras que en el canal Q se introduce un retardo T que en conjunto con el conformador provee una secuencia dada por

$$\cos \left[\pm \frac{\pi (t - T)}{2 T} \right] = \sin \left[\pm \frac{\pi t}{2 T} \right]$$

Ambas señales I y Q modulan portadoras de frecuencia f_c que se encuentran en cuadratura. También se muestra en la figura V.11b la configuración del demodulador MSK, donde se compensa en el canal I el retardo introducido en el transmisor en el canal Q . Además se ha indicado la recuperación de la portadora que, como veremos, es imprescindible por tratarse de una detección sincrónica, y la recuperación de la temporización. Un convertidor paralelo a serie vuelve a entrelazar el flujo de datos.

La amplitud y fase de las señales de salida de los moduladores es tal que su suma tiene una envolvente de amplitud constante, lo que permite su transmisión a través de dispositivos saturables.

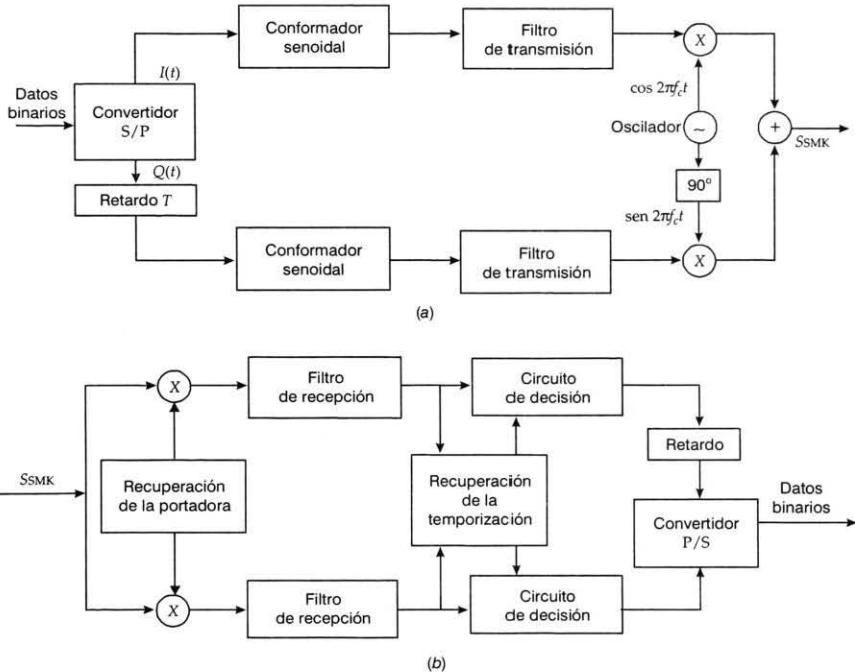


Figura V.11. Modulación MSK. a) Transmisor, b) receptor

Otra característica importante de la modulación MSK es que las transiciones de fase son lineales y continuas. Puede demostrarse que la densidad espectral de energía está dada por

$$G(f) = \frac{4 A_c^2 \pi T [1 + \cos 4 \pi T (f - f_c)]}{[\pi^2 - 16 \pi^2 T^2 (f - f_c)^2]}$$

donde A es la amplitud de la portadora.

Las señales MSK pueden generarse al igual que FSK, mediante un oscilador cuyas frecuencias alternan entre los dos valores requeridos; sin embargo, el modulador MSK difiere en que la exactitud de la desviación de frecuencia debe ser mayor.

Por supuesto que esto es imposible de lograr con un VCO, por lo que debe recurrirse a un modulador del tipo digital. Otra forma posible es basándose en la generación de señales PSK, que veremos más adelante en este capítulo.

Para detectar señales MSK es necesario reconocer la diferencia de fase de 0° o 180° entre una señal y otra, para lo cual debe utilizarse un detector coherente. En este sentido, como luego veremos, el receptor es similar a los utilizados en PSK.

V.3.4. Modulación en fase PSK

PSK, también denominada modulación discreta de fase, es una técnica de modulación digital sumamente eficiente, y muy utilizada en sistemas como enlaces satélites o radio enlaces de banda ancha.

En este tipo de modulación la información se codifica en la fase de una portadora de amplitud constante. Cuando esa información está representada por el valor absoluto de la fase, es decir referida a una portadora sin modular, se tiene el sistema PSK convencional; si la información está contenida en las variaciones de fase, es decir referida a la fase del estado anterior, tenemos los denominados sistemas diferenciales.

La ecuación que caracteriza la modulación PSK convencional para el caso binario está dada por

$$x_c(t) = x(t) \cdot A \cos(2\pi f_c t) \quad (V.16)$$

donde $x(t)$ es una señal binaria, aleatoria, de periodo T , NRZ, que toma valores $+1$ o -1 . Comparando con la expresión V.2 notamos que la única diferencia entre ASK y PSK es que en la primera la portadora se conmuta entre A y 0 , mientras que en la segunda entre $+A$ y $-A$. Las señales correspondientes a cada estado son

$$s_0(t) = A \cos(\omega_c t)$$

$$s_1(t) = -A \cos(\omega_c t) = A \cos(\omega_c t + \pi) \quad (V.17)$$

en los que se observa una diferencia de fase de 180° . Por ello esta forma de modulación recibe el nombre, exclusivamente para el caso binario, de PRK (*phase reversal keying*), aunque nosotros preferimos 2PSK. Debe observarse también que el par de ecuaciones (V.17) es antipodal, por lo que se puede predecir un buen desempeño del receptor y en consecuencia una baja probabilidad de error.

En la figura V.12 ilustramos un ejemplo de modulación PSK. Las discontinuidades de fase que aparecen al comienzo y final de cada intervalo T cuando existen transiciones de 0 a 1 o de 1 a 0 , en la práctica se atenúan debido al efecto de suavizado que en el transmisor efectúa el filtro de conformación.

La densidad espectral de energía puede calcularse transformando la ecuación (V.16).

$$G_{xc}(f) = \frac{A^2}{4} [G_x(f - f_c) + G_x(f + f_c)] \quad (V.18)$$

suponiendo que $x(t)$ es una onda binaria aleatoria con

$$G_x(f) = T \text{sinc}^2(fT) \quad (V.19)$$

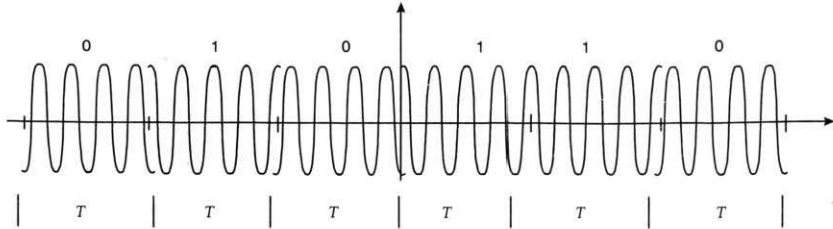


Figura V.12. Modulación PSK

resulta

$$G_{sc}(f) = \frac{A^2 T}{4} \left[\text{sinc}^2(f - f_c) T + \text{sinc}^2(f + f_c) T \right] \quad (\text{V.20})$$

La señal PSK tiene, entonces, la misma característica de doble banda lateral que la transmisión OOK, con la importante excepción de un impulso en la frecuencia de portadora. La ausencia de una componente discreta en la portadora significa que PSK posee una mejor eficiencia de potencia que OOK, aunque igual eficiencia espectral.

Si introducimos la conformación de los pulsos modulados de la ecuación (V.17) obtenemos un espectro centrado en la frecuencia portadora con un ancho de banda que duplica el mismo espectro conformado en banda base. Consecuentemente los requerimientos de ancho de banda de una señal PSK son los mismos que una ASK, a pesar de que este último proceso de modulación es lineal y el primero, en general, no lo es. Como veremos más adelante, al considerar el comportamiento en presencia de ruido, PSK exhibe un mejor desempeño que ASK y que FSK.

La gráfica de la ecuación (V.20) es la clásica función sinc, cuyo lóbulo principal tiene una anchura que es el doble de la tasa de señalización ($2/T$). Además, los lóbulos secundarios decrecen muy lentamente, por lo que es necesario filtrar la señal de salida para evitar interferencia en los sistemas adyacentes. Esto reduce la potencia transmitida, por lo que normalmente debe incrementarse la potencia del transmisor a efecto de mantener la calidad del sistema. El principal problema en el diseño de este filtro de transmisión es proveer la atenuación deseada sin causar excesiva interferencia entre símbolos (ISI). Por lo regular el mínimo ancho de banda aceptable es $2/T$, que permita la transmisión sin atenuación del lóbulo principal.

Básicamente el modulador 2PSK consiste en un proceso de conmutación que, controlada por la señal de datos, conmuta entre la portadora y su versión invertida.

En general, la modulación puede hacerse en frecuencia intermedia o bien directamente en RF. La modulación en FI puede hacerse con el conocido modulador balanceado, donde según la polaridad que asigne la señal de datos a un puente de diodos, tenemos variaciones de 180° en la fase de la señal de salida.

La modulación directamente en RF suele utilizarse cuando se desean diseños más económicos como en el caso de los enlaces de baja capacidad (2 Mbits/s y 8 Mbits/s). En la figura V.13 vemos una modulación sencilla.

La señal portadora se inyecta en el circulator 1 pasando al terminal, donde se encuentra el diodo comandado por la señal de datos 2. La atenuación de 1 a 2 es baja, del orden de décimas de dB. De igual magnitud es la atenuación cuando la señal transita en el sentido de la flecha, es decir de 2 a 3 y de 3 a 1. Por el contrario, en el sentido inverso a la flecha la atenuación es de 20 a 30 dB.

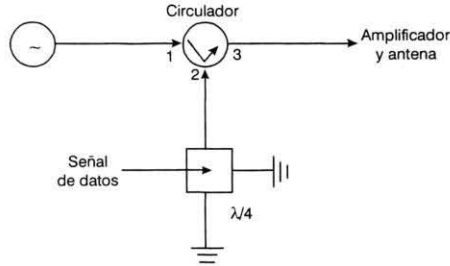


Figura V.13. Modulación directa con circulator

La señal que emerge de 2 se propaga por la guía para reflejarse en el diodo si éste está en cortocircuito, o lo atraviesa si se encuentra en circuito abierto. En este caso la señal de RF recorre una distancia adicional de $2 \lambda/4 = \lambda/2$; es decir que según la polarización del diodo comandada por la señal de datos producimos variaciones de fase de 180° , o sea estamos generando 2PSK.

El tipo de diodo se elige según la tasa de señalización requerida, teniendo en cuenta que con diodos túnel pueden lograrse tiempos de conmutación de 75 ps . En lo referente al amplificador de potencia del transmisor podemos acotar que no requiere ser lineal, por lo que se simplifica su realización. El diagrama a bloques de un receptor PSK se representa en la figura V.14.

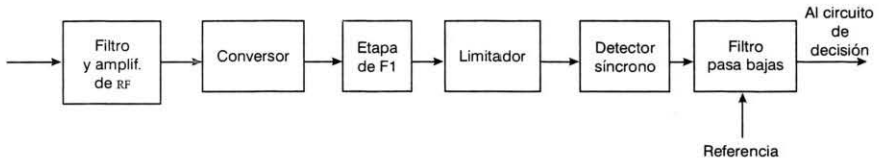


Figura V.14. Diagrama a bloques de un receptor PSK

La etapa fundamental en el receptor la constituye el detector sincrónico, donde se multiplica la señal modulada con la portadora (referencia); el producto que se obtiene se hace pasar por un filtro pasa bajas. Esquemáticamente en la figura V.15 vemos este proceso.

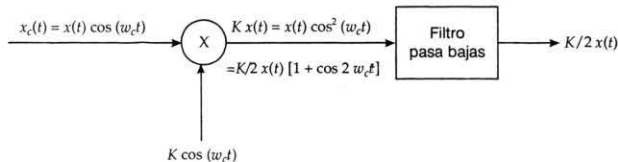


Figura V.15. Esquema de un detector sincrónico

Observe que el resultado se aplica no sólo a 2PSK, donde $x(t) = +1$ o -1 , sino a ASK, donde $x(t) = +1$ o 0 .

Hasta aquí hemos supuesto que la referencia local se encontraba en la misma frecuencia y tenía

la misma fase que la portadora de la señal recibida. Si la señal de referencia tuviese una frecuencia $w_c + \Delta w$ la salida del multiplicador sería

$$\frac{K}{2} x(t) [\Delta w \cdot t + \cos(2\Psi + \Delta w)t]$$

que una vez filtrada daría

$$\frac{K}{2} x(t) \cos(\Delta w \cdot t) \quad (V.21)$$

algo inaceptable, pues difiere totalmente de la señal deseada.

Si en cambio la señal local se encontrara en la frecuencia correcta pero presentara un corrimiento de fase θ , es decir $K \cdot \cos(w_c t + \theta)$, la señal que obtendríamos a la salida del filtro pasa bajas sería

$$\frac{K}{2} x(t) \cos(\theta) \quad (V.22)$$

o sea la señal deseada pero atenuada en amplitud. Note usted que para $\theta = \pi/2$ la salida aproximadamente es nula, para $\theta > \pi/2$, invertida; es decir, los unos transformados en ceros y los ceros en unos.

Ahora comprendemos el nombre de detección sincrónica debido a que la señal de referencia no sólo debe coincidir en frecuencia con la portadora sino también debe estar sincronizada en fase. El problema de obtener esa señal de referencia se resuelve de dos formas:

La primera consiste en enviar una señal piloto junto con la señal modulada. Dicho piloto se extrae del receptor y se utiliza para sincronizar el oscilador local. Se trata de que el piloto esté libre de ruido o, más propiamente, de que la relación señal de piloto/ruido sea aceptable.

La otra opción se basa en obtener la señal de referencia de la misma señal modulada, ya sea mediante operaciones alineales (cuadráticas) o mediante lazos de regeneración (PLL). Así, por ejemplo, si la señal modulada de entrada se rectifica en onda completa y posteriormente se filtra obtendremos una senoide de frecuencia doble a la portadora, insensible a los cambios de fase de la modulación. Esto es como consecuencia de que al duplicarse la frecuencia las fases relativas de 0° y 180° se transforman en 0° y 360° para resultar congruentes. En la figura V.16 se esquematiza este método.

La frecuencia de salida debe dividirse a la mitad a fin de tener la referencia buscada. Es obvio que la señal de salida será más depurada cuanto más agudo sea el proceso de filtrado, lo que contribuye a disminuir el jitter. En contraposición, un filtro más selectivo aumenta el tiempo de establecimiento.

V.3.5. PSK Diferencial (DPSK)

El esquema de señalización denominado *PSK diferencial* o de *comparación de la fase* evita la necesidad de obtener una referencia coherente en el receptor. En este método la referencia de fase se toma del intervalo de señalización precedente, con lo que el detector extrae la información digital, basándose en las diferencias relativas de fase; es decir, que en vez de recuperar la portadora en el receptor utilizamos la misma señal recibida pero demorada en un bit ($T = 1/r$).

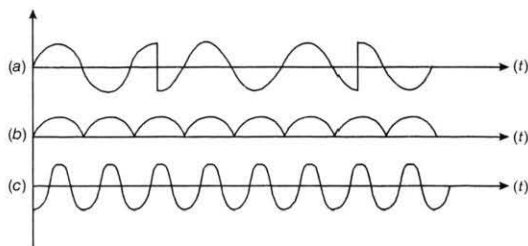


Figura V.16. Recuperación de la portadora mediante rectificación. a) Señal 2PSK (modulada), b) señal rectificada, c) señal filtrada

Por supuesto que para ello es necesario que previamente en el transmisor se codifique de manera diferencial la información digital. Los diagramas a bloques del modulador y demodulador se muestran en la figura V.17a y V.17b respectivamente, denominándose DE2PSK (*differential encoded 2PSK*).

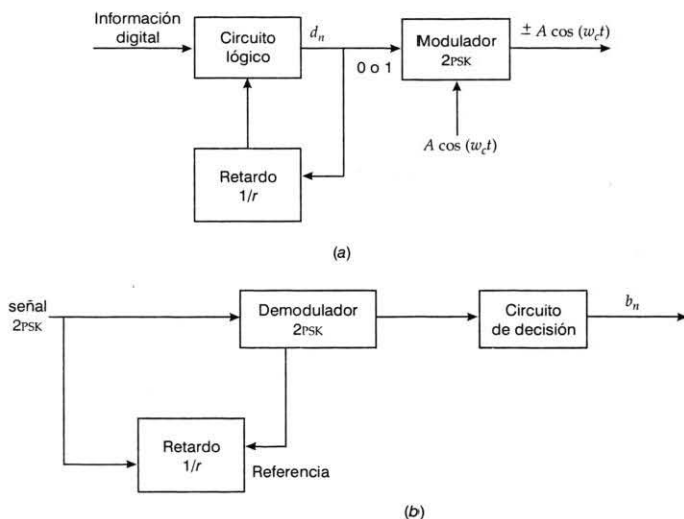


Figura V.17. a) Modulador DPSK, b) demodulador DPSK

El proceso de codificación se inicia con un bit arbitrario y a partir de ahí el flujo de datos se genera mediante

$$d_n = d_{n-1} \cdot b_n \oplus \overline{d_{n-1}} \cdot \overline{b_n} \quad (\text{V.23})$$

Las operaciones de codificación y decodificación se indican en el ejemplo del cuadro V.1.

CUADRO V.1. Codificación y decodificación diferencial

Entrada b_n	1 1 0 1 0 0 0 1
Mensaje codificado	1(#) 1 1 0 0 1 0 1 1
Fase transmitida	0 0 0 π π 0 π 0 0
Salida de la comparación de fases	+ + - + - - - +
Salida	1 1 0 1 0 0 0 1

Referencia arbitraria

El receptor DPSK correlaciona la señal recibida con una versión retardada en 1 bit de la misma señal. Su salida se compara con cero, por lo que en el convertidor A/D se opta por 1 o 0. En el cuadro V.1 podemos verificar que el receptor recupera correctamente la secuencia de entrada (en ausencia de ruido) a partir del análisis de los cambios de fase en intervalos sucesivos. El lector también puede comprobar que tomando un cero como referencia arbitraria se obtiene la misma salida.

Como veremos luego, el comportamiento de este esquema de señalización en presencia de ruido es bastante aceptable. La única desventaja significativa es que debe trabajar a velocidad fija T debido a que el retardo que se introduce en el receptor también está fijo en $1/T$; de esta manera no acepta la transmisión asíncrona.

Otra desventaja del sistema es que los errores tienden a ocurrir en pares. En efecto, si sucede un error también causa un error de decodificación en el bit siguiente. Si bien este hecho introduce una inevitable degradación ésta puede compensarse aumentando ligeramente la potencia de la señal recibida.

Antes de concluir este apartado, es oportuno comentar acerca de la utilización de la codificación diferencial, que no sólo se utiliza acompañando los sistemas de detección DPSK sino para evitar ambigüedades de fases en PSK; en efecto, tal como hemos visto en el apartado anterior, algunos métodos para obtener la portadora de referencia se basan en la duplicación de fase, por lo que existirá una incertidumbre de 180° sobre la misma. En otras palabras, obtenemos una perfecta portadora de referencia en el receptor, sincronizada en frecuencia y fase con la del transmisor, pero que puede estar desfasado en 180° , con lo que no sabemos si los unos que recibimos son en realidad ceros. Una forma de salvar esta ambigüedad consiste en transmitir una secuencia conocida cada vez que el sistema recupera sincronismo.

Como alternativa se aplica la codificación diferencial, con lo que la información ya no está contenida en la fase absoluta sino en las transiciones, y así obviar el problema de las ambigüedades.

En el receptor, como alternativa al de la figura V.17b, puede utilizarse un receptor PSK convencional (figura V.18), con su correspondiente circuito de recuperación de portadora, pero adicionándole un decodificador de lógica inversa a la del transmisor. En la figura V.18 se indica la modulación codificada diferencialmente con las dos alternativas de demodulación: la convencional, seguida del decodificador diferencial, y la demodulación diferencial.

V.4. MODULACIÓN MULTISÍMBOLO

En los sistemas de banda base se señaló que a efecto de reducir el ancho de banda puede recurrirse a esquemas de señalización de más de dos niveles.

Recordemos que idealmente (Nyquist) se pueden llegar a transmitir 2 bits/seg/Hz. Si se utiliza un conjunto de $M = 2^n$ símbolos, siendo n el número de dígitos binarios sucesivos que se combinan, pueden transmitirse $2n$ bits/seg/Hz utilizando una banda de Nyquist. Consecuentemente la señalización multinivel puede extenderse a las diversas técnicas de modulación consideradas.

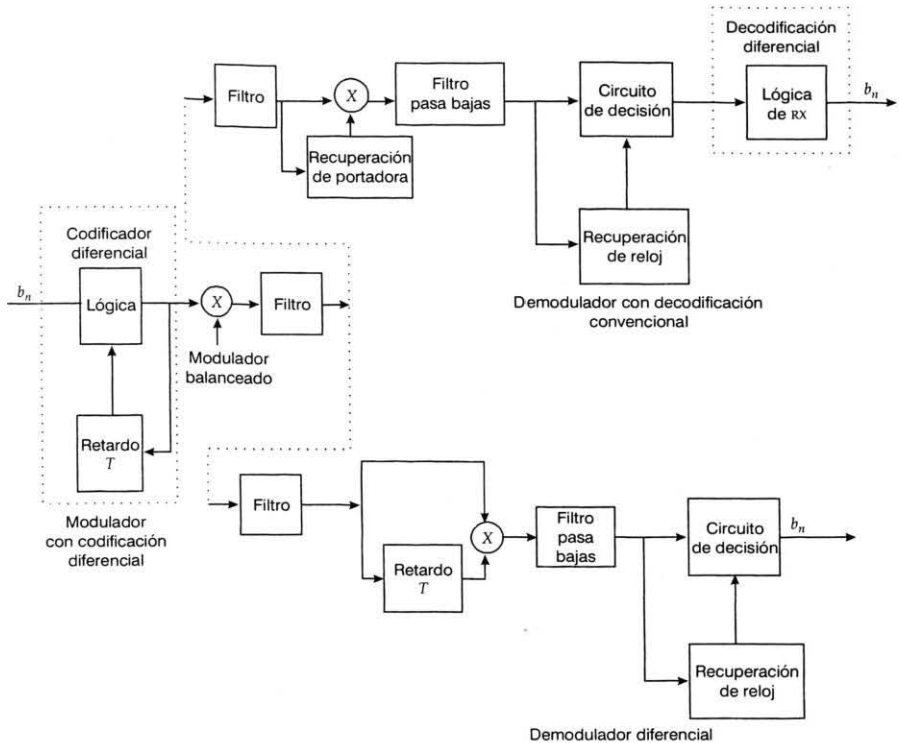


Figura V.18. Modulador codificado diferencialmente con dos opciones de demodulación

Así, tendremos modulación multisímbolo (o multinaria) en ASK, PSK y FSK. Por lo regular se prefiere utilizar esquemas multinivel cuando se quiere conservar el ancho de banda a expensas de un aumento en la potencia transmitida.

En general, las técnicas de multifase y multinivel, así como su combinación, se aplican para reducir anchos de banda y son de gran aplicación en sistemas telefónicos y satelitales. Los esquemas multifrecuencias, por el contrario, suelen producir mayores ancho de banda y como contrapartida una mejor inmunidad frente al ruido.

Para el estudio de los diversos esquemas de modulación de nivel M supondremos que el modulator toma bloques de n dígitos binarios y asigna una de las $M = 2^n$ formas de onda posibles a cada una de las diferentes M combinaciones de los n dígitos binarios.

V.4.1. PSK multisímbolo M-PSK

En el esquema M-PSK la fase de la portadora puede tomar uno de los M valores posibles separados en un ángulo $\Delta\phi = 2\pi/M$; en consecuencia las M formas de onda posibles de transmitirse están dadas por la siguiente expresión general.

$$s(t) = A \cos \left(w_c t + \frac{x(t) \Delta \phi}{2} \right) \quad (\text{V.24})$$

donde $x(t)$ es una señal de banda base, simétrica, NRZ, cuyos M niveles son: $\pm 1, \pm 3 \dots$

Consideremos primero en detalle el caso $M = 4$ conocido como 4PSK o Q-PSK (donde la Q corresponde a *Quaternary*). En este caso se combinan los dígitos binarios, pudiendo ser 00, 01, 10 y 11, con lo que deben existir cuatro ángulos de fase que les correspondan, siendo $\Delta \phi = \pi/2$

$$\frac{x(t)}{2} = \begin{cases} +\pi/4 \\ +3\pi/4 \\ -3\pi/4 \\ -\pi/4 \end{cases} \quad (\text{V.25})$$

Debe quedar claro que estos dos pulsos binarios sucesivos se almacenan para luego emitir el símbolo (la forma de onda) correspondiente. Entonces si la tasa de señalización es r bits/s cada pulso binario tendrá una duración $1/r$, pero el símbolo producido por el modulador durará $2/r$.

Si se expande la ecuación (V.24) para el caso $M = 4$ tenemos

$$s(t) = a_x \cos(w_c t) + b_x \sin(w_c t) \quad (\text{V.26})$$

donde

$$a_x = \cos \left(\frac{x(t) \Delta \phi}{2} \right) \text{ y } b_x = -\sin \left(\frac{x(t) \Delta \phi}{2} \right)$$

En consecuencia, obtenemos los siguientes pares ordenados

$$a_x, b_x = \begin{cases} (1, 1)/\sqrt{2} \\ (-1, 1)/\sqrt{2} \\ (-1, -1)/\sqrt{2} \\ (1, -1)/\sqrt{2} \end{cases} \quad (\text{V.27})$$

Observe con atención que la ecuación (V.26) corresponde a dos componentes en cuadratura, que se transmiten en forma simultánea. Tales componentes sólo podrán tomar dos valores diferentes según el flujo de datos modulante.

Estas componentes reciben la denominación de componentes en fase por el coseno, y en cuadratura por el seno, debido a su posición referida a la portadora sin modular. En realidad este tipo de expansión coincide con la que se hace para el caso de señales analógicas.

Así, por ejemplo, si para la combinación de dígitos binarios de entrada 1,1 asignamos un ángulo de fase de la portadora $\pi/4$ tendremos que la expresión (V.26) nos queda

$$s(t) = \cos(w_c t) + \sin(w_c t)$$

Es muy ilustrativo indicar las cuatro posibles posiciones que ocupa la portadora modulada recurriendo a la representación en el espacio de señal. Por claridad se representan las señales por un punto, que no es más que el extremo del vector que le corresponde en dicho espacio. En la figura V.19 se muestran los conjuntos de puntos mencionados, cuyo nombre es constelación; para mayor

claridad, se indica también el esquema 2PSK y las formas de onda correspondientes. Observe que el sentido asignado a los ejes corresponde a que en el diagrama los vectores rotan en el sentido antihorario y a que el seno adelanta $\pi/2$ al coseno.

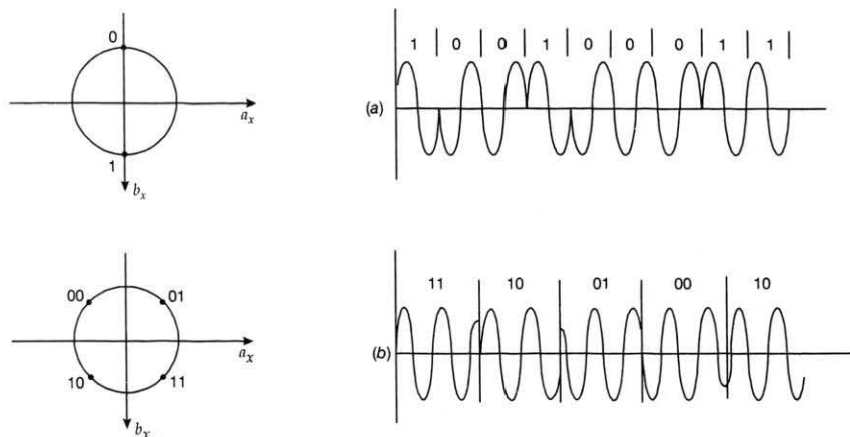


Figura V.19. Constelación y formas de onda para a) PSK, b) 4PSK

En el cuadro V.2 se indica el valor de las componentes en cuadratura en función de los dígitos binarios de entrada. Se muestra también la expresión correspondiente de la portadora modulada (ecuación V.24).

CUADRO V.2. PSK Componentes en cuadratura

Dígitos binarios	Coeficientes		Portadora modulada
	a_x	b_x	
01	0.707	-0.707	$\cos(w_c t + \pi/4)$
00	-0.707	-0.707	$\cos(w_c t + 3\pi/4)$
10	-0.707	0.707	$\cos(w_c t - 3\pi/4)$
11	0.707	0.707	$\cos(w_c t - \pi/4)$

La modulación 4PSK se hace mediante diversos métodos: el primero de ellos se basa en la combinación lineal de dos señales en cuadratura según se desprende de la ecuación (V.26), tal como se muestra en la figura V.20.

En la misma figura se almacena un par de bits de la información digital en un registro de desplazamiento; y se aplican a dos moduladores PSK binarios cuyas portadoras están en cuadratura entre sí. Observe que cada modulador recibe un flujo de datos a la mitad de la velocidad con la que ingresa la información digital. Las dos señales PSK binarias se combinan —sumadas— para producir los cuatro estados posibles que se reseñan en el cuadro V.2.

Modulación digital

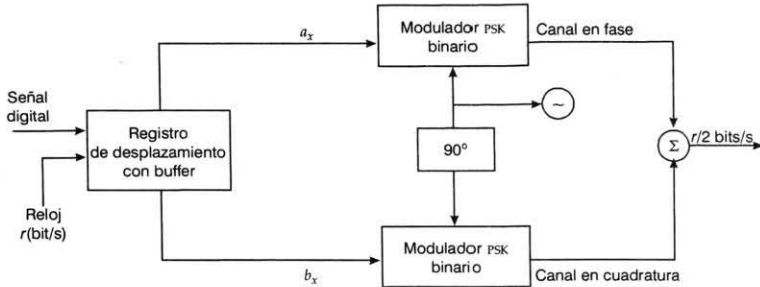


Figura V.20. Modulador 4PSK

En lo que respecta al proceso de detección, ya no nos basta con una sola referencia sincrónica que nos provea

$$y_{ax} = \cos \left(w_c t + \frac{x(t) \Delta \phi}{2} \right) \cos(w_c t)$$

que convenientemente filtrada nos dará

$$y_{Ax} = \cos \left(\frac{x(t) \Delta \phi}{2} \right)$$

Ahora será necesaria otra referencia ortogonal

$$y_{bx} = \cos \left(w_c t + \frac{x(t) \Delta \phi}{2} \right) \sin(w_c t)$$

que después del filtro será:

$$y_{Bx} = -\sin \left(\frac{x(t) \Delta \phi}{2} \right)$$

De esta manera todas las decisiones del detector pueden basarse en la polaridad de y_{Ax} y y_{Bx} . En efecto, observando el cuadro V.2 resulta que el primer bit es cero (0) cuando el ángulo de fase es positivo ($\pi/4$ o $3\pi/4$); es decir que este primer bit está completamente especificado por la polaridad de $\sin [x(t) \Delta \phi]/2$ que es precisamente la salida y_{Bx} del receptor sincrónico con referencia ortogonal. Idéntico razonamiento puede emplearse para el segundo bit que será uno cuando la fase sea $\pm\pi/4$, esto es, cuando el coseno sea positivo. En la figura V.21 vemos la realización básica de un detector 4PSK, así como la etapa de regeneración. Observe que de acuerdo con lo expuesto, consiste en dos canales, cada uno de ellos operando en forma binaria; cada canal posee un detector de fase seguido de un filtro pasa bajas. La salida de los mismos se aplica a un regenerador y también al circuito de recuperación de reloj (*clock*). Posteriormente el flujo de datos regenerado pasa por un decodificador diferencial que elimina la decodificación hecha en el transmisor; finalmente los dos canales se multiplexan bit a bit compensando el efecto que produce el registro de desplazamiento del modulador.

Considerando las figuras V.20 y V.21 se evidencia un concepto importante: la información digital de entrada se ha dividido en dos secuencias binarias intercaladas de velocidad $r/2$, las cuales

modulan en PSK binaria dos componentes en cuadratura. Estas secuencias binarias están recién recompuestas en el multiplexor de salida del detector; de esta manera, para efecto del análisis, la modulación 4PSK puede verse como dos 2PSK en cuadratura. Una consecuencia inmediata de lo expuesto es la necesidad de una perfecta ortogonalidad entre las portadoras de los dos canales, pues en caso contrario se produciría una suerte de interferencia diafónica. La otra consecuencia es que, como toda modulación en cuadratura, inhibe el uso de banda lateral única, ya que la remoción de parte de la banda introduce componentes en el plano ortogonal a la misma.

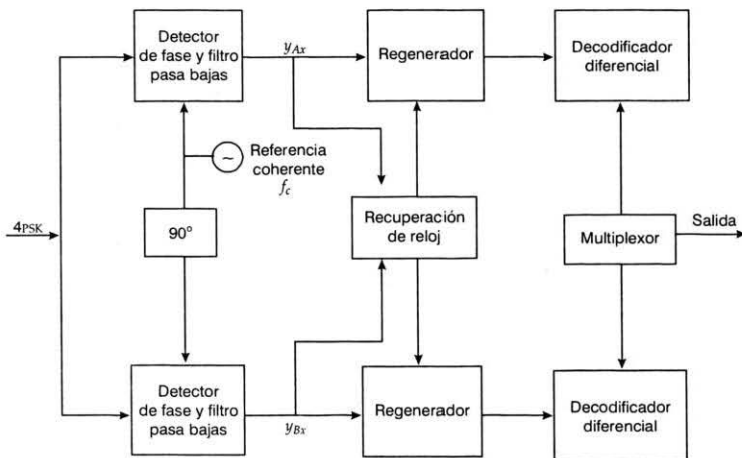


Figura V.21. Detector 4PSK

Ahora consideraremos $M = 8$ denominado 8PSK. En este caso las ocho fases diferentes están separadas en un ángulo $\Delta \varphi = \pi/4$. Resulta evidente entonces que cada fase representa un grupo de tres dígitos binarios (tribits). En el cuadro V.3 se indica la expresión de la portadora correspondiente a las ocho posibles combinaciones de tres dígitos binarios, además del valor de las componentes en cuadratura.

CUADRO V.3. 8PSK Componentes en cuadratura

Dígitos binarios	Coeficientes		Portadora modulada
	a_x	b_x	
011	0.924	-0.383	$\cos(w_c t + \pi/8)$
010	0.383	-0.924	$\cos(w_c t + 3\pi/8)$
000	-0.383	-0.924	$\cos(w_c t + 5\pi/8)$
001	-0.924	-0.383	$\cos(w_c t + 7\pi/8)$
101	-0.924	0.383	$\cos(w_c t - 7\pi/8)$
100	-0.383	0.924	$\cos(w_c t - 5\pi/8)$
110	0.383	0.924	$\cos(w_c t - 3\pi/8)$
111	0.924	0.383	$\cos(w_c t - \pi/8)$

En la figura V.22 se muestra la constelación correspondiente a 8PSK. Observe que los tribits de cada fase siguen el código Gray a fin de minimizar la severidad de los posibles errores.

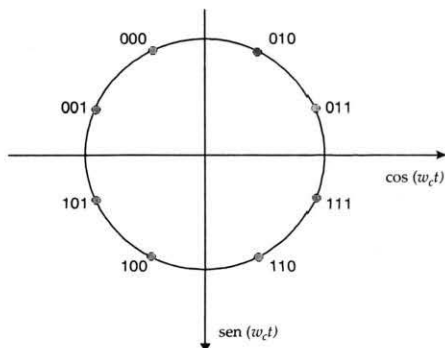


Figura V.22. Diagrama de fase de una señal 8PSK

Para generar señales 8PSK pueden utilizarse los métodos expuestos para 4PSK. Cabe aclarar, sin embargo, que ahora las componentes en cuadratura ya no sólo varían en signo sino en amplitud, por lo que de querer utilizarse un esquema similar al de la figura V.20 tendría que transformarse el registro de desplazamiento en un generador de señales de banda base, para que entregara a los moduladores las magnitudes de a_x y b_x , según los valores del tribit que se almacene en un intervalo dado.

Un ejemplo de modulador 8PSK básico lo vemos en la figura V.23.

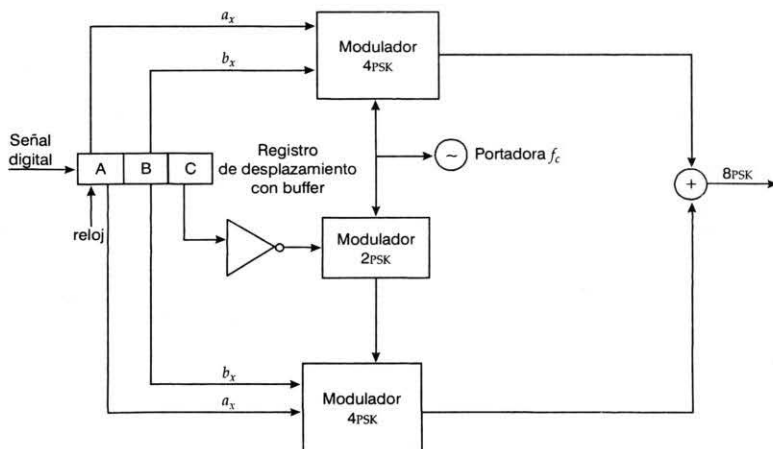


Figura V.23. Modulador 8PSK

El mismo modular 8PSK consta de dos moduladores 4PSK y un modulador 2PSK. La salida del modulador 4PSK superior está dada por los dos primeros bits (*A* y *B*) del tritbit que se está transmitiendo. Lo mismo sucede con el modulador 4PSK inferior si el tercer bit (*c*) es un 0; pero si el tercer bit es un 1 su constelación se modifica tal como se indica en la figura. La composición de ambas salidas en el sumador, finalmente nos entrega la constelación 8PSK deseada.

En lo referente a la detección de señales 8PSK vemos que si aplicamos un circuito como el de la figura V.21 con tan sólo dos referencias (*sen* y *cos*), no podremos discriminar entre puntos de la constelación de un mismo cuadrante. La salida de los detectores de fase ya no es constante y tendríamos que introducir una referencia para discriminar su amplitud. Obviamente esto no es deseable pues preferimos mantener una detección basada en polaridades. Para solucionar estos inconvenientes existen dos métodos posibles.

El primer método consiste en introducir dos referencias adicionales en el receptor y medir la fase de la señal recibida también respecto a ellas. En la figura V.24 se indica la posición de las nuevas fases de referencia denominadas *c* y *d*, de las cuales son $\cos(w_c t + \pi/4)$ y $\sin(w_c t + \pi/4)$, respectivamente, así como las ya usadas *a* y *b*.

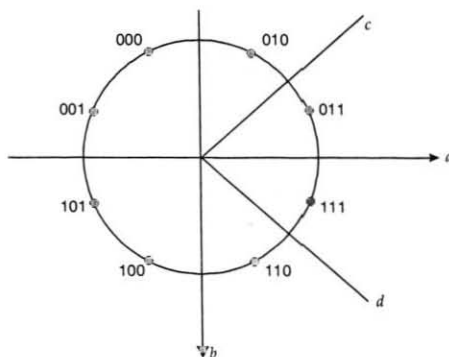


Figura V.24. Fases de referencia para la detección 8PSK

Las salidas de los cuatro detectores se determinan por

$$\begin{aligned}
 y_{ax} &= \cos \left[w_c t + \frac{x(t) \Delta \varphi}{2} \right] \cos(w_c t) \\
 y_{bx} &= \cos \left[w_c t + \frac{x(t) \Delta \varphi}{2} \right] \sin(w_c t) \\
 y_{cx} &= \cos \left[w_c t + \frac{x(t) \Delta \varphi}{2} \right] \cos(w_c t + \pi/2) \\
 y_{dx} &= \cos \left[w_c t + \frac{x(t) \Delta \varphi}{2} \right] \sin(w_c t + \pi/2)
 \end{aligned} \tag{V.28}$$

Estas salidas convenientemente filtradas mediante un filtro pasa bajas nos entregarán, amén de una constante:

$$\begin{aligned}y_{Ax} &= \cos \left[\frac{x(t) \Delta \phi}{2} \right] \\y_{Bx} &= -\sin \left[\frac{x(t) \Delta \phi}{2} \right] \\y_{Cx} &= 0.707 \left[\cos \left(\frac{x(t) \Delta \phi}{2} \right) + \sin \left(\frac{x(t) \Delta \phi}{2} \right) \right] \\y_{Dx} &= 0.707 \left[\cos \left(\frac{x(t) \Delta \phi}{2} \right) - \sin \left(\frac{x(t) \Delta \phi}{2} \right) \right]\end{aligned}\tag{V.29}$$

Examinando la figura V.24, podemos decir que:

- El primer bit será un 1 si $y_{Bx} > 0$
- El segundo bit será un 1 si $y_{Ax} > 0$
- El tercer bit será un 1 si y_{Cx} y y_{Dx} son simultáneamente positivos o negativos.

En resumen, si llamamos A , B , C y D a las variables lógicas que representan salidas positivas de y_{Ax} , y_{Bx} , y_{Cx} y y_{Dx} , los tres bits estarán dados por:

Primer bit = B

Segundo bit = A (V.30)

Tercer bit = $CD + \overline{CD}$

El segundo método para detectar 8PSK se basa en que y_{Cx} y y_{Dx} pueden obtenerse a partir de la combinación lineal de Ax y Bx , tal como lo indican las ecuaciones (V.29). En efecto,

$$y_{Cx} = 0.707 (y_{Ax} - y_{Bx}) \tag{V.31}$$

$$y_{Dx} = 0.707 (y_{Ax} + y_{Bx})$$

Note que las ecuaciones (V.31) representan una rotación de $\pi/4$ de los vectores ortogonales base; si el ángulo de rotación fuese un valor α arbitrario tendríamos

$$\begin{aligned}y_{Cx} &= \cos(\alpha) y_{Bx} - \sin(\alpha) y_{Ax} \\y_{Dx} &= \sin(\alpha) y_{Bx} + \cos(\alpha) y_{Ax}\end{aligned}\tag{V.32}$$

lo que permite extender lo expuesto cuando son necesarias nuevas referencias, como por ejemplo en 16PSK. La realización circuital basada en la ecuación (V.31) se muestra en la figura V.25.

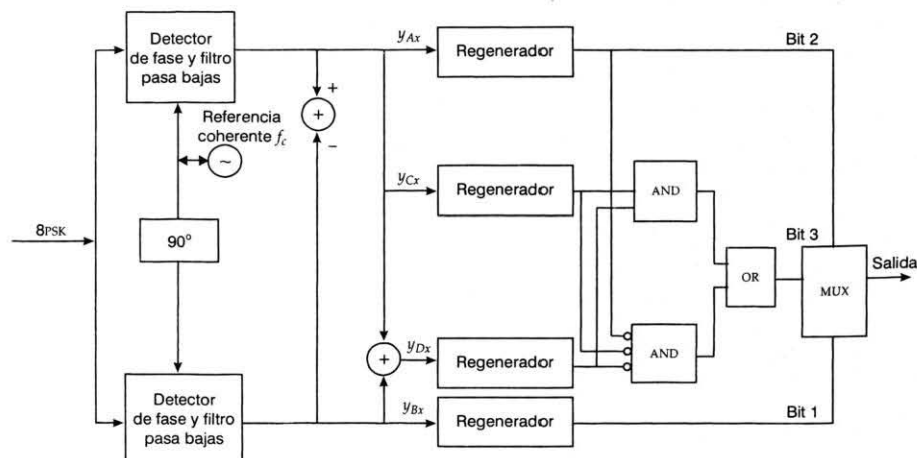


Figura V.25. Detector 8PSK de dos referencias

Observe que se ha obviado la ponderación por el coeficiente 0.707, pues nos basta con las polaridades obtenidas.

Espectro de las señales PSK

En páginas anteriores analizamos el espectro de la señal 2PSK; ahora bien, una señal 4PSK de velocidad r bits/s puede considerarse como dos canales 2PSK en cuadratura, de velocidad $r/2$ bits/s. Debido a la ortogonalidad de los dos canales las señales no están correlacionadas, por lo que el espectro compuesto es meramente la suma de los espectros individuales; en consecuencia, y como cabría esperar, el ancho de banda requerido para 4PSK es la mitad del 2PSK a igual velocidad de señalización.

Un análisis similar basado en las señales de banda base, permite concluir que si para 8PSK la tasa de señalización se reduce a $r/3$, el espectro se reducirá en igual proporción. En conclusión, todos los sistemas PSK convencionales producen un espectro de la forma $\text{sinc}(x)$ centrada en la frecuencia de portadora. Debe tenerse en cuenta, además, que para las condiciones de equiprobabilidad y simetría de la constelación no hay componente espectral en la frecuencia de portadora.

Cabe mencionar, no obstante, que el gran ahorro se obtiene al pasar de 2PSK a 4PSK. Este hecho, sumado a ciertas características que describiremos frente al ruido, y a una moderada dificultad en la implementación, hacen de 4PSK el esquema de modulación más difundido. Sin embargo, cuando la reducción de ancho de banda es el objetivo primordial, es necesario recurrir a métodos de modulación más complejos. Resulta claro que los lóbulos laterales decrecen muy lentamente. Así, por ejemplo, la atenuación relativa del primer lóbulo lateral respecto al principal es de sólo 13.5 dB. Si bien podemos encarar diversas definiciones de ancho de banda, resulta ilustrativo aplicar la adoptada por la FCC (Federal Communications Commission), que establece que el ancho de banda ocupado es la banda que deja 0.5% de la señal por encima y por debajo de los límites superiores e inferiores respectivamente; es decir que 99% de la potencia de la señal debe estar dentro de la banda ocupada.

De lo expuesto concluimos que la gran dispersión de su espectro es uno de los principales inconvenientes de los sistemas de modulación PSK y debe tomarse en cuenta en el diseño de los filtros de transmisión. Así, también los organismos reguladores de la utilización espectral fijan especificaciones estrictas tendientes a evitar interferencias entre sistemas.

V.4.2. Modulación por cuadratura de amplitud QAM

Hasta ahora en los esquemas PSK todos los puntos de la constelación se encontraban sobre una circunferencia, lo que implicaba amplitud constante. Y si bien habíamos visto la conveniencia de suponer dos canales en cuadratura, los niveles de las señales modulantes (banda base) en cada canal no eran independientes pues la composición de ambos debía resultar en una señal de amplitud constante.

Si ahora abandonamos esta condición y permitimos que las señales de banda base en los dos canales en cuadratura sean totalmente independientes, estamos en presencia de un esquema denominado QAM (*Quadrature Amplitude Modulation*) o bien APK (*Amplitud Phase Keying*).

Dicho esquema consta de la modulación multinivel de amplitud de dos portadoras en cuadratura en forma independiente; en consecuencia, los dos canales en cuadratura son completamente independientes, incluso la codificación en banda base.

La ecuación de la señal QAM puede escribirse como una extensión de la (V.26)

$$s_i(t) = a_i \cos(w_c t) + b_i \sin(w_c t) \quad (\text{V.33})$$

donde a_i y b_i toman en forma independiente los valores discretos previstos según el número de niveles establecidos, siendo $M = L^2$.

Para el caso especial de dos niveles (± 1) en cada canal el sistema no es más que 4PSK. En la figura V.26 se representa el esquema donde cada canal en cuadratura puede tomar cuatro niveles distintos, cuyo resultado es el denominado 16QAM, es decir $L = 4$ y $M = 16$.

Observe que a diferencia de 16PSK la envuelta de la portadora modulada ya no será constante, lo que en principio inhibe el uso de dispositivos saturables.

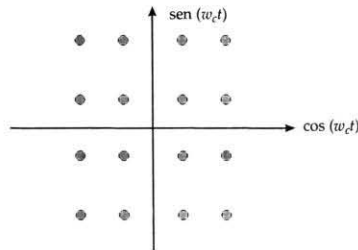


Figura V.26. Diagrama de constelación de 16QAM

Por otro lado, la distancia entre puntos en un esquema QAM es siempre mayor que un esquema PSK equivalente, lo que puede verificarse a partir de $d = 2 \sin(\pi/M)$ y $d = \sqrt{2}/(M^{1/2} - 1)$, que dan la distancia entre puntos para los esquemas PSK y QAM respectivamente. Como veremos más adelante,

esta conclusión tiene una importancia fundamental en el rendimiento de ambos sistemas en lo que se refiere al ruido.

El diagrama a bloques de un modulador QAM está representado en la figura V.27. El flujo de datos de entrada, con una tasa de señalización de r bits/s, se divide en dos flujos de velocidad $r/2$. A continuación un convertidor de 2 a L niveles transforma una señal binaria en una señal multinivel cuya velocidad es

$$r_s = \frac{r}{2} \cdot \frac{1}{\log_2 L} \text{ (símbolos /s)}$$

Así, por ejemplo, si la fuente de datos tiene una velocidad de r 10 Mbits/s, el flujo dividido será de 5 Mbits/s y si usamos un sistema 16QAM tendremos $L = 4$ niveles en cada señal a_i y b_i , con lo que sus velocidades serán

$$r_s = \frac{10 \text{ Mbits} / 2}{2} \cdot \frac{1}{\log_2 4} = 2.5 \text{ Mbits} / \text{s}$$

Como veremos más adelante, el ancho de banda final requerido dependerá del filtro de conformación ubicado a la salida de los convertidores de 2 a L niveles.

Una limitación importante de este método de modulación es la necesidad de utilizar amplificadores de potencia muy lineales con mínimos efectos de conversión AM a AM y AM a PM.

Sin embargo, para mejorar las relaciones señal-ruido en los sistemas satelitales es altamente deseable la generación de potencias de salida altas, lo que se logra más fácilmente utilizando dispositivos limitados en amplitud. Ejemplos de estos elementos son los tubos de ondas progresivas, los klystrons en saturación, los amplificadores transistorizados operados en clase C, así como los amplificadores basados en diodos Gunn e Impatt (ILA).

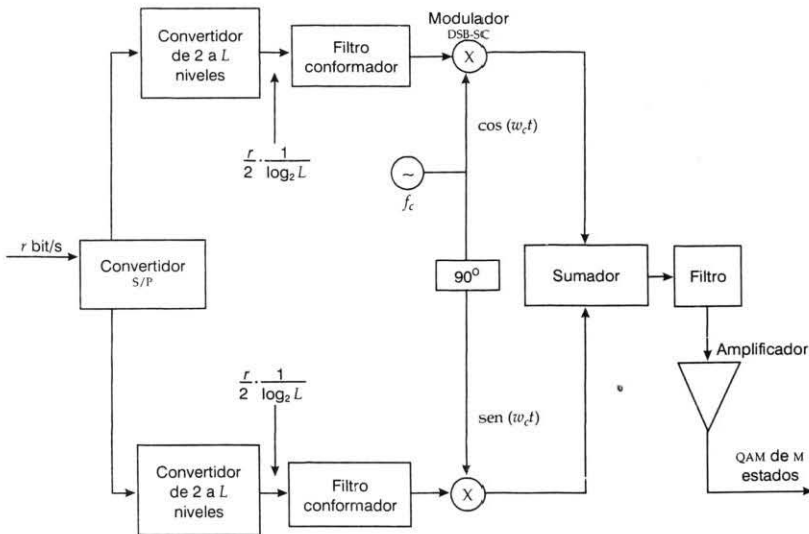


Figura V.27. Diagrama a bloques de un modulador QAM de M estados ($M = L^2$).
(dsb-sc doble banda lateral con portadora suprimida)

Un método para compatibilizar los requerimientos de linealidad y alta potencia de salida consiste en obtener la señal 16QAM como combinación de dos señales 4PSK. Puesto que las señales 4PSK no se filtran antes de combinarse no contienen variación de amplitud y en consecuencia no sufren ninguna degradación si se amplifican en un dispositivo a lineal. La tensión de salida de uno de los amplificadores se hace al doble de la del otro amplificador, y de la combinación lineal de ambos se obtiene la señal 16QAM, la cual es filtrada para limitar su espectro.

El diagrama a bloques de un demodulador QAM de $M = L^2$ niveles se aprecia en la figura V.28; el convertidor de L a 2 niveles consiste en $L - 1$ comparadores que prevén un 1, lógico a su salida si en el instante de muestreo la señal de entrada supera el nivel de umbral preestablecido. En caso contrario se tendrá un 0 lógico. Las salidas de los $L - 1$ comparadores se conectan a un circuito lógico que finalmente determinará la palabra binaria de salida correspondiente; así, para $M = 16$ se tendrá $L = 4$ y un dabit de salida por cada instante de muestreo del convertidor L a 2.

El muestreo se hace a $r/(2 \log_2 L)$ y su tasa de salida binaria es $r/2$. Finalmente el convertidor paralelo-serie multiplexa los dos flujos de datos proveyendo la salida deseada a r bits/seg.

El espectro de una señal QAM se determina, como en todos los moduladores de producto, a partir del espectro de las señales de banda base aplicadas a los canales en cuadratura.

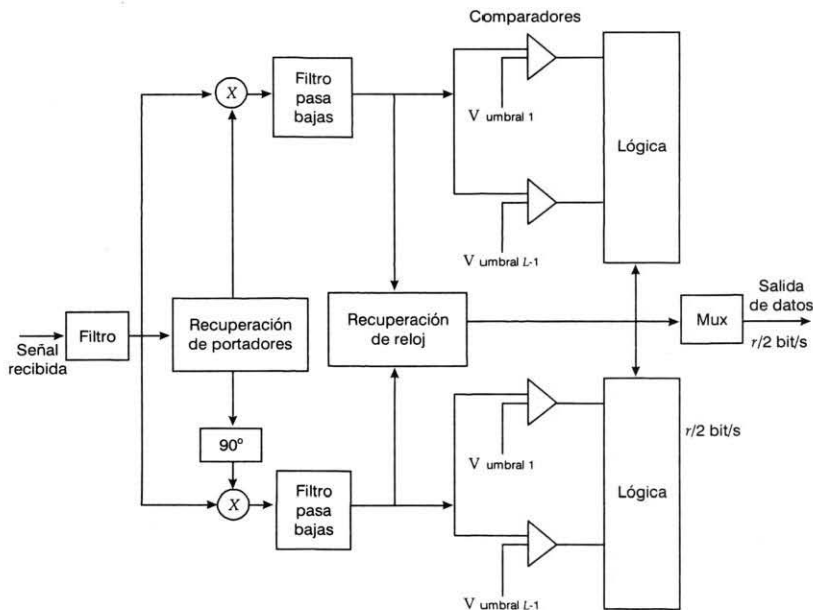


Figura V.28. Diagrama a bloques de un demodulador QAM de $M = L^2$ niveles

Como estas señales tienen estructuras similares a los PSK, los espectros de QAM son similares a los PSK de igual orden. Concretamente el espectro de una señal 16QAM es similar a un 16PSK, y así sucesivamente.

En conclusión, las señales deben filtrarse para su transmisión; pero la limitación en banda debe

hacerse cuidadosamente para evitar interferencias entre símbolos (ISI). En la práctica se recurre a conformar la señal con caídas (*roll-off*) sinusoidal similares a las descritas en la transmisión en banda base. Si bien la conformación puede efectuarse directamente sobre los pulsos antes de la modulación, en la práctica se hace tanto en el transmisor como en el receptor, acercándonos a una característica de Nyquist.

Recordando que en un canal de banda base de B (Hz) puede transmitirse con una velocidad

$$r = \frac{2B}{1 + \alpha} \quad (\text{V.34})$$

donde el factor de *roll off* α varía de 0 para un filtro pasa bajas ideal a 1 para un filtro de coseno elevado. Ahora bien, para un canal de radio acoplado a un modulador lineal (del tipo de los descritos) necesitamos un ancho de banda $B_T = 2B$ con lo que la velocidad de señalización será

$$r = \frac{B_T}{1 + \alpha} \quad (\text{V.35})$$

Finalmente, para un esquema QAM con $M = 2^n$ estados posibles la tasa de señalización es de

$$r = \frac{n B_T}{1 + \alpha} \quad (\text{V.36a})$$

lo que equivale a

$$\delta = \frac{n}{1 + \alpha} \text{ bits/seg/Hz} \quad (\text{V.36b})$$

En el cuadro V.4 aparecen valores típicos de la ecuación (V.36b).

CUADRO V.4. Densidad de información en QAM (bits/seg/Hz)

M	Factor de roll-off				
	0.1	0.2	0.5	0.75	1
2	0.9	0.83	0.67	0.57	0.5
4	1.8	1.66	1.33	1.14	1.0
8	2.7	2.5	2.0	1.71	1.5
16	3.6	3.32	2.67	2.28	2.0

Ejemplo V.1

Calcule la tasa de señalización posible en un canal telefónico cuyo ancho de banda se estima en 2.4 KHz. El tipo de modulación usado es 4QAM (equivalente a 4PSK) y se suponen factores de *roll-off* de 10% y 100 por ciento.

Para $\alpha = 10\%$ la tasa de señalización se aplica la ecuación (V.36)

$$r = \frac{2 \cdot 2.4 \text{ KHz}}{1 + 0.1} = 4\,300 \text{ bits/s}$$

y para $\alpha = 100\%$ se tiene

$$r = \frac{2 \cdot 2.4 \text{ KHz}}{1 + 1} = 2\,400 \text{ bits/s}$$

Ejemplo V.2

El *Spade* es un sistema de comunicaciones satelital de acceso múltiple que utiliza un esquema de modulación 4PSK para transmitir 64 Kbits/s en un ancho de banda de 38 KHz. En consecuencia, el factor de *roll-off* necesario será

$$\alpha = \frac{n B_t}{r} - 1 = 0.19 = 19\%$$

V.4.3. Esquemas FSK de M estados

La técnica FSK de M estados mantiene la potencia constante a expensas de incrementar el ancho de banda. Si consideramos un esquema de M estados tendremos que las M posibles señales utilizadas están definidas por

$$s_i(t) = \begin{cases} A \cos(w_i t) & \text{para } 0 \leq t \leq T \\ 0 & \text{en otro caso} \end{cases} \quad (\text{V.37})$$

y la condición de ortogonalidad

$$\int_0^T s_i(t) s_j(t) dt = \begin{cases} \frac{A^2 T}{2} & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases} \quad (\text{V.38})$$

Es decir que las señales tienen una duración T e igual energía. Si suponemos que se utiliza conformación de Nyquist y que las señales están espaciadas en $1/T$ Hz, el ancho de banda total estará dado por M/T .

Los problemas involucrados en el cálculo del espectro de MFSK son similares a los de FSK binaria, por lo que pueden calcularse espectros típicos con el uso de computadoras. En general el aspecto

de ambos espectros coincide. El receptor óptimo para este conjunto ortogonal de señales consiste en un banco de M filtros acoplados como se indica en la figura V.29.

La salida de los filtros se muestra en los instantes kT y se decide cuál $s_i(t)$ estaba presente en ese intervalo.

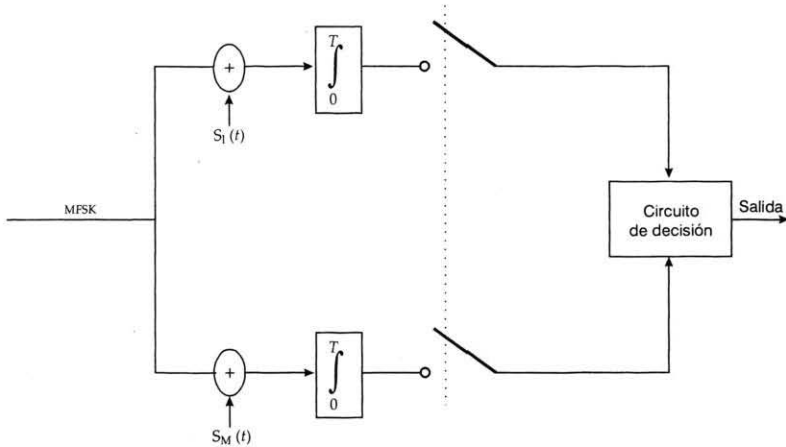


Figura V.29. Receptor de MFSK

V.5. COMPORTAMIENTO EN PRESENCIA DE RUIDO

En el desarrollo anterior presentamos las diversas técnicas de modulación digital; describimos distintos aspectos de los extremos transmisor y receptor y dimos especial atención a cada ancho de banda de transmisión. Esto último es coherente con lo señalado en el capítulo I respecto a considerar el ancho de banda como una de las limitaciones fundamentales de un sistema de comunicaciones. En concreto, el conocimiento de los anchos de banda requeridos implica mantener la distorsión y, en consecuencia, la interferencia entre símbolos por debajo de límites aceptables. También apuntamos en el capítulo I que el ruido constituye otra de las limitaciones fundamentales, pues combinado con la atenuación de la señal en su trayecto transmisor-receptor, conspira contra la eficiente transmisión de la información.

En este apartado consideraremos el comportamiento de los diversos sistemas de modulación digital en presencia de ruido. Supondremos, al igual que en banda base, que el ruido se adiciona a la señal mientras ésta se propaga por el sistema; esto es, ruido aditivo, y de naturaleza gaussiana como se muestra en la figura V.30. Esta suposición, aparte de ser realista, nos permitirá utilizar modelos simples, así como adecuados parámetros de evaluación y comparación.

En el caso de los esquemas de modulación binarios ASK, PSK, FSK, notamos que pueden generalizarse con la expresión:

$$S(t) = \begin{cases} f_1(t) & \text{si } b_K = 0 \\ f_2(t) & \text{si } b_K = 1 \end{cases}$$

Modulación digital

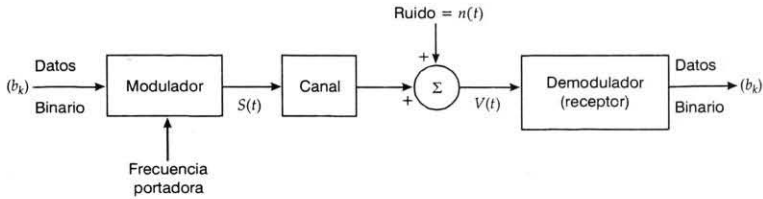


Figura V.30. Sistema de transmisión de datos binarios en presencia de ruido

V.5.1. Filtro acoplado

El receptor tiene que determinar cuál de las dos formas de onda conocidas $f_1(t)$ o $f_2(t)$ estuvo presente durante cada intervalo de la señal. En el caso de transmisión binaria, los mensajes se transmiten únicamente por dos símbolos; uno de ellos puede representarse por un pulso $s(t)$, y el otro por la ausencia del pulso, como se muestra en forma general en la figura V.31.

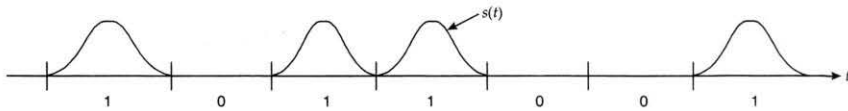


Figura V.31. Concepción general de los esquemas de modulación binarios

Estamos entonces en el caso de banda base, y podemos aplicar los mismos resultados que obtuvimos en el capítulo IV para la detección en presencia de ruido; es decir, debemos muestrear los pulsos recibidos y compararlos contra un nivel de decisión para detectar cuál es la forma de onda que ha sido transmitida.

Es conveniente apuntar que en los receptores, previo a la etapa de detección, siempre existirá un filtro, que se denomina filtro acoplado, destinado a reducir la relación señal a ruido en el instante de decisión, tal como se muestra en la figura V.32.

El problema real no es detectar la forma de onda de la señal sino determinar cuándo el pulso está presente o ausente. El detector óptimo debe ser aquel que tenga la mínima probabilidad de error al efectuar la decisión.

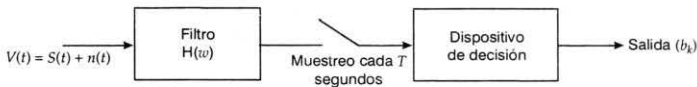


Figura V.32. Diagrama a bloques general de un receptor coherente

El proceso de decisión puede facilitarse si la señal pasa a través de un filtro que acentúe la señal de interés $s(t)$ y suprima el ruido $n(t)$. Esto equivale a maximizar la razón de la amplitud de la señal a la razón del ruido, en algún instante $t = t_m$.

Es más conveniente usar un filtro que maximice el cuadrado de las amplitudes, es decir, hay que maximizar la razón

$$\rho = \frac{s_0^2(t)}{n_0^2(t)} \quad (\text{V.39})$$

donde

$s_0(t)$ es la componente de la señal a la salida del filtro

$\overline{n_0^2(t)}$ es el valor cuadrático medio del ruido a la salida del filtro

Estas relaciones se ilustran en la figura V.33.

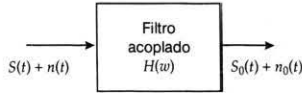


Figura V.33. Relaciones entrada-salida de un filtro acoplado

Para tener totalmente determinado el filtro acoplado debemos conocer su función de transferencia $H(\omega)$ o su respuesta al impulso; empezaremos por escribir la respuesta del filtro en la frecuencia (sólo se indica la parte de componente de la señal de información).

$$S_0(\omega) = H(\omega) S(\omega)$$

$$s_0(t) = \mathcal{F}^{-1} [H(\omega) S(\omega)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(\omega) S(\omega) e^{j\omega t} d\omega \quad (\text{V.40})$$

Al evaluarla en el instante de decisión

$$s_0(t_m) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(\omega) S(\omega) e^{j\omega t_m} d\omega \quad (\text{V.41})$$

Suponiendo que

$G(\omega) \equiv$ es el espectro de densidad de potencia de ruido $n(t)$ a la entrada del filtro.
con lo que

$|H(\omega)|^2 G(\omega) \equiv$ es el espectro de densidad de potencia $n_0(t)$ a la salida del filtro; luego, el valor cuadrático medio del ruido a la salida del filtro está dado por:

$$\overline{n_0^2(t)} = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega) |H(\omega)|^2 d\omega \quad (\text{V.42})$$

Evalutando en el instante de decisión $t = t_m$

$$\overline{n_0^2(t_m)} = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega) |H(\omega)|^2 d\omega \quad (\text{V.43})$$

Suponiendo que el ruido en el canal es blanco con una densidad de potencia $G(\omega) = \frac{N}{2}$, tenemos que

$$\overline{n_0^2(t)} = \frac{N}{4\pi} \int_{-\infty}^{\infty} |H(\omega)|^2 d\omega \quad (\text{V.44})$$

sustituyendo las expresiones (V.44) y (V.41) en (V.39) se obtiene

$$\rho = \frac{S_0^2(t_m)}{n_0^2(t_m)} = \frac{\frac{1}{4\pi^2} \left| \int_{-\infty}^{\infty} H(\omega) S(\omega) e^{j\omega t_m} d\omega \right|^2}{\frac{N}{4\pi} \int_{-\infty}^{\infty} |H(\omega)|^2 d\omega} \quad (\text{V.45})$$

Si utilizamos la desigualdad de Schwartz que establece que

$$\left| \int_{-\infty}^{\infty} F_1(\omega) F_2(\omega) d\omega \right|^2 \leq \int_{-\infty}^{\infty} |F_1(\omega)|^2 d\omega \cdot \int_{-\infty}^{\infty} |F_2(\omega)|^2 d\omega$$

donde la igualdad ocurre si $F_1(\omega) = K F_2^*(\omega)$; $K = \text{cte. arbitraria}$

Haciendo $F_1(\omega) = H(\omega)$ y $F_2(\omega) = S(\omega) e^{j\omega t_m}$ obtenemos

$$\left| \int_{-\infty}^{\infty} H(\omega) S(\omega) e^{j\omega t_m} d\omega \right|^2 \leq \int_{-\infty}^{\infty} |H(\omega)|^2 d\omega \cdot \int_{-\infty}^{\infty} |S(\omega)|^2 d\omega \quad (\text{V.46})$$

Al sustituir las ecuaciones (V.46) en (V.45) obtenemos

$$\rho = \frac{S_0^2(t_m)}{n_0^2(t_m)} \leq \frac{1}{N\pi} \frac{\int_{-\infty}^{\infty} |H(\omega)|^2 d\omega \cdot \int_{-\infty}^{\infty} |S(\omega)|^2 d\omega}{\int_{-\infty}^{\infty} |H(\omega)|^2 d\omega} = \frac{1}{N\pi} \int_{-\infty}^{\infty} |S(\omega)|^2 d\omega$$

$$\rho_{\max} = \frac{S_0^2(t_m)}{n_0^2(t_m)} = \frac{1}{N\pi} \int_{-\infty}^{\infty} |S(\omega)|^2 d\omega \quad (\text{V.47})$$

lo que ocurre únicamente cuando la desigualdad se torna en igualdad, y esto es posible cuando

$$H(\omega) = K S^*(\omega) e^{-j\omega t_m} = K S(-\omega) e^{-j\omega t_m} \quad (\text{V.48})$$

donde K es una constante arbitraria.

Entonces, la respuesta al impulso $h(t)$ del filtro óptimo está dada por

$$h(t) = F^{-1}[H(\omega)] = F^{-1}[K S(-\omega) e^{-j\omega t_m}] \quad (\text{V.49})$$

Como $S(-\omega) \longleftrightarrow s(-t)$ y el término $e^{-j\omega t_m}$ representa un desplazamiento de t_m segundos, con lo que tenemos

$$h(t) = s(t_m - t) \quad (\text{V.50})$$

es decir, la respuesta al impulso del filtro óptimo es igual a la señal de entrada invertida y desplazada por t_m segundos. Donde se hizo por conveniencia $K = 1$.

Lo que falta por determinar es el instante de decisión t_m ; para determinarlo graficaremos la ecuación (V.50); en la figura V.34a se muestra un pulso arbitrario de duración T ; en la V.34b se ilustra el pulso invertido tal como indica la ecuación (V.50). Finalmente, en c , d y e vemos la respuesta del filtro acoplado para este pulso con desplazamientos $t_m < T$, $t_m = T$ y $t_m > T$ respectivamente.

La respuesta de la figura V.34c representa un sistema no causal (el pulso contiene tiempos negativos) que como sabemos no es físicamente realizable; en otras palabras, en este caso el filtro debería dar una respuesta cuando todavía no terminamos de excitarlo. La figura V.34d corresponde al caso donde el pulso ha terminado de excitar completamente el filtro acoplado; finalmente en el inciso e el desplazamiento es mayor a la duración del pulso, y aunque no aparece en la figura, en este caso la salida del filtro también se ve afectada por el pulso precedente.

Con base en la explicación anterior es claro que el caso óptimo que debemos seleccionar es $t_m = T$ que corresponde al caso causal y la respuesta del filtro acoplado sólo se debe a un único pulso. Luego, la respuesta del filtro acoplado debe ser,

$$h(t) = s(T - t) \quad (\text{V.51})$$

En la figura V.35 se muestra un diagrama a bloques del filtro acoplado indicando su respuesta al impulso.

Calculemos ahora el valor de la componente de la señal a la salida $s_0(t)$ y el ruido a la salida $n_0^2(t)$ del receptor acoplado.

La energía de la señal $s(t)$ está dada por:

$$E = \int_{-\infty}^{\infty} s^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |S(\omega)|^2 d\omega$$

entonces

$$\rho = \frac{S_0^2(t_m)}{n_0^2(t_m)} = \frac{1}{N\pi} \int_{-\infty}^{\infty} |S(\omega)|^2 d\omega = \frac{2E}{N} \quad (\text{V.52})$$

donde $N/2$ es la densidad de potencia del ruido. La amplitud de la señal $s_0(t_m)$ se obtiene sustituyendo la ecuación (V.48) en la (V.41):

$$s_0(t_m) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(\omega) S(\omega) e^{j\omega t_m} d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(-\omega) S(\omega) e^{j\omega t_m} e^{-j\omega t_m} d\omega$$

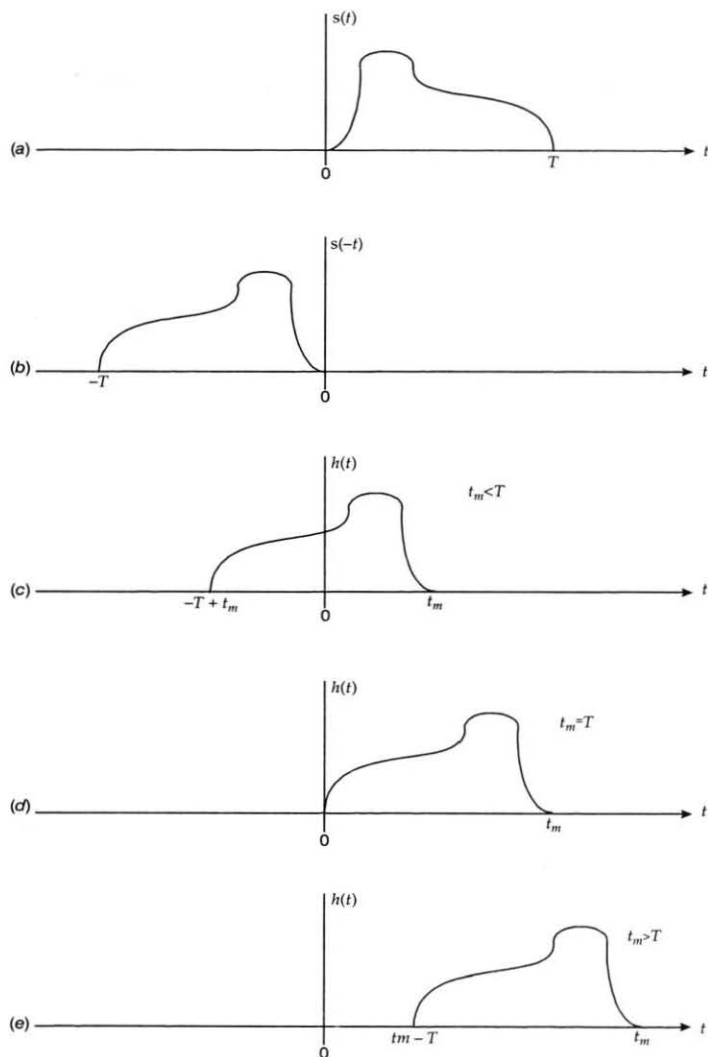


Figura V.34. Respuesta al impulso del filtro acoplado para varios valores de desplazamiento

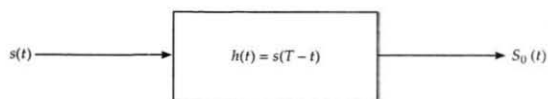


Figura V.35. Filtro acoplado

$$s_0(t_m) = \frac{1}{2\pi} \int_{-\infty}^{\infty} |S(\omega)|^2 d\omega = E \quad (\text{V.53})$$

La máxima amplitud de la componente de la señal a la salida ocurre en $t = t_m$ y tiene una magnitud igual a la energía E de la señal $s(t)$.

El resultado anterior es importante ya que nos dice que la máxima amplitud de la señal de salida es independiente a la forma de onda de entrada $s(t)$ y únicamente depende de su energía; este resultado se ilustra en la figura V.36.

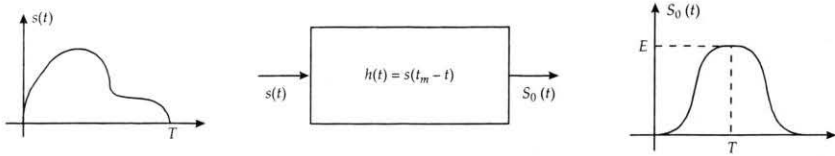


Figura V.36. Ilustración de la ecuación V.53

El valor cuadrático medio del ruido a la salida se obtiene sustituyendo la ecuación (V.53) en (V.52):

$$\frac{s_0^2(t_m)}{n_0^2(t_m)} = \frac{2E}{N} \quad (\text{V.54})$$

$$\frac{E^2}{n_0^2(t_m)} = \frac{2E}{N}$$

$$\therefore \overline{n_0^2(t)} = \frac{NE}{2} \quad (\text{V.55})$$

El filtro igualador puede hacerse por un arreglo alternativo. Si la entrada al filtro igualador es $f(t)$, entonces su salida $r(t)$ está dada por

$$r(t) = \int_{-\infty}^{\infty} f(x) h(t-x) dx$$

donde $h(t)$ es la respuesta al impulso del filtro igualador y está dada por

$$h(t) = s(t_m - t) \Rightarrow h(t-x) = s(t_m - t + x)$$

$$r(t) = \int_{-\infty}^{\infty} f(x) s(t_m - t + x) dx$$

como la decisión a la salida se hace en $t = t_m$

$$r(t_m) = \int_{-\infty}^{\infty} f(x) s(x) dx \quad (V.56)$$

Un diagrama a bloques de la ecuación (V.56) se muestra en la figura V.37.

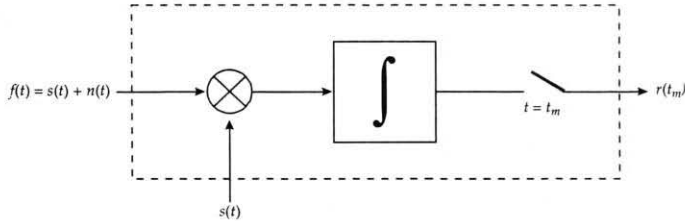


Figura V.37. Receptor de correlación o detector coherente (síncrono), ya que el receptor necesita una réplica en fase de la señal de entrada

Detección asíncrona (detector de envolvente)

Otro método de detección muy utilizado es el denominado detector de envolvente, que evita los problemas de temporización y la fase de detección coherente. En este caso la señal de entrada de alta frecuencia pasa a través de un dispositivo no lineal y un filtro pasa bajas. Una forma común del detector de envolvente es un diodo rectificador de media onda (dispositivo no lineal) seguido por un filtro RC (figura V.38). Como su nombre lo indica, la salida del detector representa la envolvente de la señal de alta frecuencia. La constante de tiempo RC debe seleccionarse para retener la amplitud de la señal de entrada para varios periodos de la señal portadora. Cabe hacer notar, sin embargo, una dificultad de este esquema: como la señal PSK tiene una envolvente constante, el detector de envolvente no puede utilizarse para demodular una señal PSK; entonces forzosamente la modulación PSK requiere una detección coherente.

V.5.2. Detección coherente de señales binarias

Comenzaremos este apartado con el estudio de las señales ASK, FSK y PSK detectadas coherentemente. Comparando las expresiones de estas señales dadas por las ecuaciones (V.2), (V.6) y (V.17), respectivamente, vemos que pueden generalizarse en la expresión

$$x(t) = A x(t) \cos(\omega_c t) \quad (V.57)$$

donde $x(t)$ para:

ASK variará entre $+A$ y 0

PSK variará entre $-A$ y A

FSK será siempre igual a A pero ω_c tomará valores ω_1 o ω_2 .

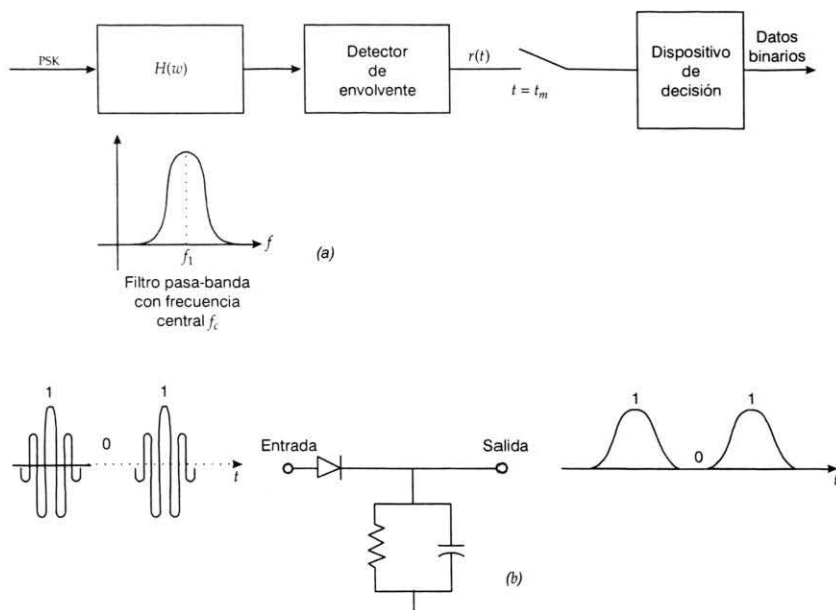


Figura V.38. a) Diagrama a bloques del receptor de envolvente, b) detector de envolvente

Ahora bien, si suponemos que la probabilidad de los unos es igual a la probabilidad de los ceros, la probabilidad de error será igual a la probabilidad de error encontrada para los sistemas de banda base, que es (véase capítulo IV):

$$P_e = \text{Erfc} \left[\frac{A_S}{2\sigma} \right] \quad (\text{V.58})$$

en este caso la amplitud A_S debe tomarse a la salida del filtro acoplado, pero como vimos en la sección anterior la máxima amplitud del pulso ocurre cuando el instante de muestreo $t_m = T$ y es igual a la energía de $x(t)$ (en este caso), por lo que a partir de (V.57) tenemos

$$A_S = E = A^2 T / 2 \quad (\text{V.59})$$

donde T es la duración de un símbolo. Como ASK puede compararse con el caso de pulsos polares, el nivel de decisión óptimo es:

$$a_{\text{opt ASK}} = \frac{A_S}{2} = \frac{E}{2} = \frac{A^2 T}{4} \quad (\text{V.60})$$

utilizando los resultados del filtro acoplado

$$\sigma^2 = \overline{n_b^2(t)} = \frac{N E}{2} \quad (\text{V.61})$$

donde $N/2$ es la densidad de potencia del ruido. Sustituyendo estas ecuaciones en la (V.58) obtenemos la probabilidad de error para ASK en caso de detección coherente.

$$P_{e\text{ASK}} = \text{Erfc} \left[\sqrt{\frac{E}{2n}} \right] = \text{Erfc} \left[\frac{A}{2} \sqrt{\frac{T}{N}} \right] \quad (\text{V.62})$$

donde:

E = energía de $x(t)$.

T = duración del pulso.

$N/2$ = densidad de potencia del ruido.

A = amplitud pico de la señal modulada.

También se puede expresar la probabilidad de error en términos de la potencia promedio de la señal modulada. Como en este caso (ASK) $x(t)$ está presente la mitad del tiempo en promedio y para el resto del tiempo no hay señal, entonces

$$P_s = \frac{1}{T} \cdot \frac{E}{2} = \frac{A^2}{4} \quad (\text{V.63})$$

$$P_{e\text{ASK}} = \text{Erfc} \left[\sqrt{\frac{P_s T}{N}} \right] \quad (\text{V.64})$$

De manera similar se obtiene la probabilidad de error para la modulación FSK (suponiendo señales bipolares véase figura V.9).

$$P_{e\text{ASK}} = \text{Erfc} \left[\sqrt{\frac{E}{N}} \right] = \text{Erfc} \left[A \sqrt{\frac{T}{2N}} \right] \quad (\text{V.65})$$

donde:

E = energía de la señal bipolar = $A^2 T/2$.

T = duración del pulso.

$N/2$ = densidad de potencia del ruido.

A = amplitud pico de la señal modulada.

$\sigma^2 = \overline{n_b^2(t)} = NE = NA^2 T/2$

en este caso el nivel de decisión es

$$a_{\text{opt. FSK}} = E = \frac{A^2 T}{2} \quad (\text{V.66})$$

La potencia promedio para FSK está dada por

$$P_s = \frac{1}{T} \cdot \frac{E}{2} = \frac{A^2}{2} \quad (\text{V.67})$$

$$P_{e\text{ASK}} = \text{Erfc} \left[\sqrt{\frac{P_s T}{N}} \right] \quad (\text{V.68})$$

de donde se deduce que la probabilidad de error expresada en términos de la potencia promedio es igual tanto en ASK como en FSK.

Finalmente, para la modulación PSK su probabilidad de error está dada por

$$P_{e\text{ASK}} = \text{Erfc} \left[\sqrt{\frac{2E}{N}} \right] = \text{Erfc} \left[A \sqrt{\frac{T}{2N}} \right] \quad (\text{5.69})$$

donde:

E = energía de la señal modulada = $A^2 T/2$.

T = duración del pulso.

$N/2$ = densidad de potencia del ruido.

A = amplitud pico de la señal modulada.

$$\sigma^2 = \overline{n_o^2(t)} = NE/2 = NA^2 T/4$$

El nivel de decisión óptimo en este caso es $a = 0$. La potencia promedio está dada por:

$$P_s = \frac{E}{T} = \frac{A^2}{2} \quad (\text{V.70})$$

$$P_{e\text{ASK}} = \text{Erfc} \left[\sqrt{\frac{2P_s T}{N}} \right] \quad (\text{V.71})$$

En el cuadro V.5 se comparan los parámetros de ancho de banda y de probabilidad de error para los esquemas de modulación binaria ASK, FSK y PSK. Teniendo presente que la función de error complementario es una función decreciente, concluiremos que el esquema que mejores características presenta es PSK (tanto en ancho de banda como en probabilidad de error).

CUADRO V.5. Comparación de esquemas de modulación binarios

Tipo de modulación	Ancho de banda	Tipo de receptor	$Pe = \text{Erfc}[x]$
ASK	$2B$	Coherente/envolvente	$x = \sqrt{\frac{A^2 T}{4N}}$
FSK	$2B + 2\Delta f$	Coherente/envolvente	$x = \sqrt{\frac{A^2 T}{2N}}$
PSK	$2B$	Coherente	$x = \sqrt{\frac{A^2 T}{N}}$

V.5.3. Detección diferencial coherente

Para concluir el estudio de las señales binarias consideraremos el caso DPSK refiriéndonos a la figura V.17b.

En lo que respecta al comportamiento en presencia del ruido, podría suponerse que la detección diferencial requiere el doble de potencia que la detección coherente debido a que la referencia de fase también está contaminada de ruido; sin embargo, en la realidad las perturbaciones tienden a cancelarse en el proceso de comparación y en consecuencia la degradación no resulta tan elevada.

Puede demostrarse que para este caso la probabilidad de error es

$$P_e = \frac{1}{2} e^{-A^2/2\sigma^2} = \frac{1}{2} e^{-E_b/N} \quad (V.72)$$

donde E_b es la energía por bit y $N/2$ es la densidad de potencia del ruido. La ecuación anterior está representada en la figura V.39. Del análisis resulta claro que DPSK tiene una ventaja de 3 dB sobre los demás sistemas no coherentes y una desventaja menor que 1 dB respecto a PSK para $P_e = 10^{-4}$. Como contrapartida recordemos que DPSK no requiere sincronización de fase pero su implementación es más compleja que OOK y FSK no coherentes.

V.5.4. Detección de señales multinivel en presencia de ruido

La probabilidad de error de cualquier señal digital está directamente relacionada con la distancia entre los puntos del espacio de señal. Cuanto mayor es esta distancia menor será la posibilidad de que, por efecto del ruido, el receptor confunda un punto con otro.

En el caso de señales PSK multifase la característica del sistema puede evaluarse comparando la disminución de la distancia entre puntos referido al sistema 2PSK. Otro elemento para tener en cuenta es el ancho de banda relativo que requiere el esquema de modulación en estudio, ya que el mismo determina el ruido del sistema.

Veamos como ejemplo el esquema 4PSK, donde la distancia entre puntos se reduce en $1/\sqrt{2}$ (3 dB) respecto a 2PSK.

Esto supone una desmejora en el rendimiento del sistema; sin embargo, para igual velocidad de señalización, el ancho de banda de 4PSK tal como lo hemos expuesto se reduce a la mitad, por lo que el ruido a la entrada del receptor disminuye también en 3 dB. Estos dos hechos hacen que sorprendentemente 4PSK exhiba igual probabilidad de error con la mitad de ancho de banda al precio de una ligera dificultad de realización.

En general la expresión que nos da la distancia entre puntos adyacentes en M-PSK para un sistema de valor pico unidad es

$$d = 2 \sin(\pi/4) \quad (V.73)$$

Las ecuaciones que expresan la P_e en función de la relación energía por símbolo a densidad de potencia de ruido pueden verse en [1], [2], [3], [4]; así, por ejemplo, de [3] tenemos

$$P_e \cong 2Q\left[\sqrt{\frac{2E}{N}} \sin^2\left(\frac{\pi}{M}\right)\right] \quad (5.74)$$

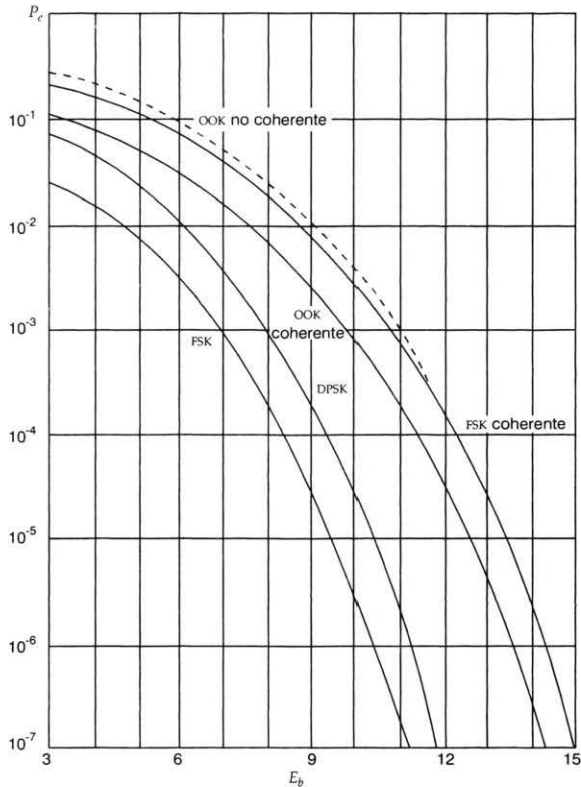


Figura V.39. Probabilidad de error para diversos sistemas binarios de modulación

En la figura V.40 se ha representado la probabilidad de error para distintos sistemas M-PSK.

Para el esquema de modulación QAM podemos aplicar un razonamiento similar para la evaluación de la probabilidad de error. En este caso la distancia entre puntos de señal adyacentes para un sistema de valor pico unidad es

$$d = \frac{\sqrt{2}}{(\sqrt{M} - 1)} \quad (\text{V.75})$$

La comparación de las ecuaciones (V.73) y (V.75) junto con la consideración de igualdad de ancho de banda ya mencionada nos indica que QAM presenta una mayor inmunidad al ruido que PSK de igual orden a igualdad de potencia pico. Si la comparación se hace en términos de potencia media la ventaja de QAM se acentúa más.

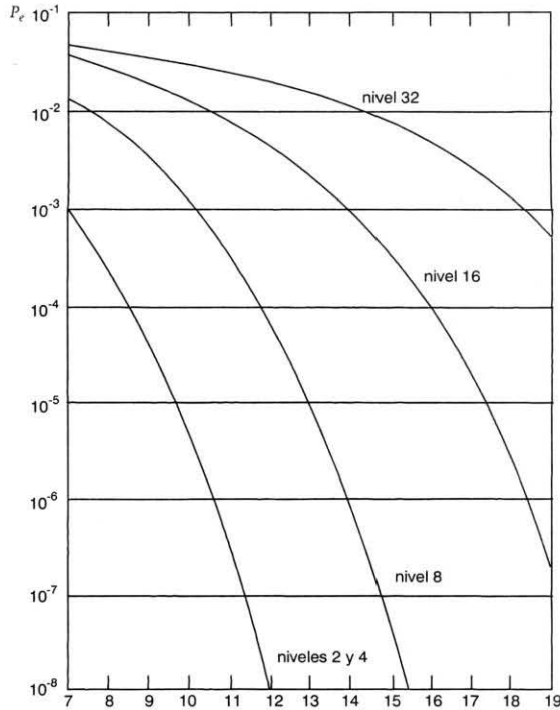


Figura V.40. Probabilidad de error M-PSK

Para una constelación QAM *cuadrada*, tal como se muestra en la figura V.26, la probabilidad de error es

$$P_e \cong 4 \left(1 - \frac{1}{\sqrt{m}} \right) Q \left[\sqrt{\frac{3E}{(M-1)N}} \right] \quad (\text{V.75})$$

Para concluir, la figura V.41 representa la probabilidad de error para diversos esquemas QAM.

V.6. RECOMENDACIONES INTERNACIONALES PARA MÓDEMS

En el cuadro V.6 presentamos las características principales de las recomendaciones de la serie V para módems de la UIT. En el cuadro V.7 se muestran las asignaciones de fase para las recomendaciones que utilizan el esquema de modulación M-PSK.

Cabe hacer notar que en ninguna recomendación se utiliza el esquema de modulación ASK por su alta probabilidad de error; asimismo los esquemas de modulación binarios se utilizan sólo para bajas velocidades; para velocidades más altas se recomiendan, en general, esquemas multinivel.

CUADRO V.6. Resumen de las recomendaciones de módems de banda de voz de la serie V

Núm.	Recomendación	Velocidad (bits/sec)	Método de transmisión	Separación de canal	Frecuencia de portadora (Hz)	Técnica de modulación
1	V.21	300	Full duplex	División de frecuencia	1080 y 1750	Corrimiento de frecuencia
2	V.22	1200	Full duplex	División de frecuencia	1200 y 2400	Corrimiento de fase
3	V.22	600	Full duplex	División de frecuencia	1200 y 2400	Corrimiento de fase
4	V.22bis	2400	Full duplex	División de frecuencia	1200 y 2400	Modulación cuadratura-amplitud
5	V.22bis	1200	Full duplex	División de frecuencia	1200 y 2400	Modulación cuadratura-amplitud
6	V.23	600	Half duplex	NA	1300 y 1700	Corrimiento de frecuencia
7	V.23	1200	Half duplex	NA	1300 y 2100	Corrimiento de frecuencia
8	V.26	2400	Full duplex	NA	1800	Corrimiento de fase
9	V.26bis	2400	Half duplex	NA	1800	Corrimiento de fase
10	V.26bis	1200	Half duplex	NA	1800	Corrimiento de fase
11	V.26ter	2400	Ambas	Cancelación de eco	1800	Corrimiento de fase
12	V.26ter	1200	Ambas	Cancelación de eco	1800	Corrimiento de fase
13	V.27	4800	Ambas	ND	1800	Corrimiento de fase
14	V.27bis	4800	Ambas	4 hilos	1800	Corrimiento de fase
15	V.27bis	2400	Ambas	4 hilos	1800	Corrimiento de fase
16	V.27ter	4800	Half duplex	Ninguna	1800	Corrimiento de fase
17	V.27ter	2400	Half duplex	Ninguna	1800	Corrimiento de fase
18	V.29	9600	Ambas	4 hilos	1700	Modulación cuadratura-amplitud
19	V.29	7200	Ambas	4 hilos	1700	Corrimiento de fase
20	V.29	4800	Ambas	4 hilos	1700	Corrimiento de fase
21	V.32	9600	Full duplex	Cancelación de eco	1800	Modulación cuadratura-amplitud
22	V.32	9600	Full duplex	Cancelación de eco	1800	Modulación trellis
23	V.32	4800	Full duplex	Cancelación de eco	1800	Modulación cuadratura-amplitud

1	V.21	300	1:1	Ambas	No
2	V.22	600	2:1	Ambas	No
3	V.22	600	1:1	Ambas	No
4	V.22bis	600	4:1	Ambas	No
5	V.22bis	600	2:1	Ambas	No
6	V.23	600	NA	Ambas	Sí
7	V.23	1200	NA	Ambas	Sí
8	V.26	1200	2:1	Síncrono	Sí
9	V.26bis	1200	2:1	Síncrono	Sí
10	V.26bis	1200	1:1	Síncrono	Sí
11	V.26ter	1200	2:1	Ambas	No
12	V.26ter	1200	1:1	Ambas	No
13	V.27	1600	3:1	Síncrono	Sí
14	V.27bis	1600	3:1	Síncrono	Sí
15	V.27bis	1200	2:1	Síncrono	Sí
16	V.27ter	1600	3:1	Síncrono	Sí
17	V.27ter	1200	2:1	Síncrono	Sí
18	V.29	2400	4:1	Síncrono	No
19	V.29	2400	3:1	Síncrono	No
20	V.29	2400	2:1	Síncrono	No
21	V.32	2400	4:1	Síncrono	No
22	V.32	2400	5:1	Síncrono	No
23	V.32	2400	2:1	Síncrono	No

CUADRO V.6 (concluye)

Núm.	Recomendación	Líneas comutadas	Líneas privadas	Uso de V.28	(dB-25) Conector ISO	Ecuatización	Alertorización	Uso de V.54
1	V.21	Sí	No	Sí	2110	ND	ND	ND
2	V.22	Sí	Punto a punto, 2 hilos	Sí	2110	Fija	Sí	Sí
3	V.22	Sí	Punto a punto, 2 hilos	Sí	2110	Fija	Sí	Sí
4	V.22bis	Sí	Punto a punto, 2 hilos	Sí	2110	Fija/adaptativa	Sí	Sí
5	V.22bis	Sí	Punto a punto, 2 hilos	Sí	2110	Fija/adaptativa	Sí	Sí
6	V.23	Sí	No	Sí	2110	ND	No	ND
7	V.23	Sí	No	Sí	2110	ND	No	ND
8	V.26	No	Punto a punto, Multipunto, 4 hilos	Sí	2110	ND	No	ND
9	V.26bis	Sí	No	Sí	2110	Fija	No	ND
10	V.26bis	Sí	No	Sí	2110	Fija	No	ND
11	V.26ter	Sí	Punto a punto, 2 hilos	Sí	2110	Ambas	Sí	Sí
12	V.26ter	Sí	Punto a punto, 2 hilos	Sí	2110	Ambas	Sí	Sí
13	V.27	No	Sí	Sí	2110	Manual	Sí	ND
14	V.27bis	No	2 hilos, 4 hilos	Sí	2110	Adaptativa	Sí	ND
15	V.27bis	No	2 hilos, 4 hilos	Sí	2110	Adaptativa	Sí	ND
16	V.27ter	Sí	No	Sí	2110	Adaptativa	Sí	ND
17	V.27ter	Sí	No	Sí	2110	Adaptativa	Sí	ND
18	V.29	No	Punto a punto, 4 hilos	Sí	2110	Adaptativa	Sí	ND
19	V.29	No	Punto a punto, 4 hilos	Sí	2110	Adaptativa	Sí	ND
20	V.29	No	Punto a punto, 4 hilos	Sí	2110	Adaptativa	Sí	ND

21	V.32	Sí	Punto a punto, 2 hilos	Sí	2110	Adaptativa	Sí	Sí
22	V.32	Sí	Punto a punto, 2 hilos	Sí	2110	Adaptativa	Sí	Sí
23	V.32	Sí	Punto a punto, 2 hilos	Sí	2110	Adaptativa	Sí	Sí

ND = No Disponible.

NA = No Aplicable.

CUADRO V.7. Asignación de fases para las recomendaciones que utilizan modulación M-PSK

Núm. recomendación	Velocidad (bits/seg)	0°	45°	90°	135°	180°	225°	270°	315°
V.22	1200	01		00		10		11	
V.22	600			0				1	
V.26	2400	00		01		11		10	
V.26bis	2400		00		01		11		10
V.26bis	1200			0				1	
V.26ter	2400	00		01		11		10	
V.26ter	1200	0				1			
V.27	4800	001	000	010	011	111	110	100	101
V.27bis	4800	001	000	010	011	111	110	100	101
V.27bis	2400	00		01		11		10	
V.27ter	4800	001	000	010	011	111	110	100	101
V.27ter	2400	00		01		11		10	

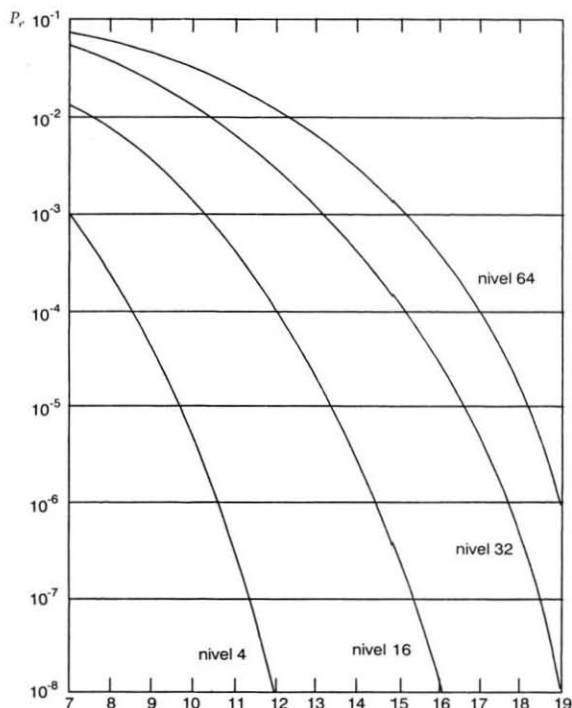


Figura V.41. Probabilidad de error en M-QAM

V.7. PROBLEMAS

- V.1. Un módem transmite a una velocidad de 4.8 Kbits/seg. Los pulsos de transmisión tienen una amplitud de 12 V y la amplitud *rms* del ruido (σ) es de 5 mV. Calcule la probabilidad de error para los siguientes casos: ASK, FSK, PSK, basándose en el análisis de los resultados anteriores; ¿qué esquema de modulación seleccionaría usted para una realización práctica?, explique sus comentarios.

NOTA: La función de error complementario puede aproximarse como

$$\text{Erfc}[x] \approx \frac{e^{-x^2}}{x\sqrt{\pi}}$$

- V.2. Se usa modulación FSK para transmitir a través de una red telefónica cuyo ancho de banda es de 500 a 2 900 Hz. Se desea un índice de modulación $\beta = 0.7$, de tal manera que las frecuencias portadoras estén a 1 200 y 2 200 Hz respectivamente. Encuentre el ancho de banda máximo de la señal digital.

- V.3. Un canal de voz se transmite usando técnicas PCM; se toman 8 000 muestras por segundo y se utiliza una cuantificación de 7 bits; la señal binaria resultante se transmite usando pulsos conformados senoidalmente con un factor de *roll-off* de 20 por ciento.
- Determine la velocidad de la señal PCM en bits/seg.
 - Determine el ancho de banda de la señal PCM.
 - Para transmisión, las señales se modulan en PSK. ¿Cuál es el ancho de banda requerido en este caso?
 - Si se modula en PSK de 4 fases, ¿cuál es el ancho de banda requerido?
 - ¿Cuál es el ancho de banda si se utiliza FSK? Si la desviación en frecuencia es de 38 kHz, ¿cuánto vale el índice de modulación β ? ¿Cuál es el espaciamiento entre las portadoras?
- V.4. La salida de un sistema PCM consiste en una secuencia binaria de pulsos que ocurre a una frecuencia de 2×10^6 bits/seg; calcule el BW de transmisión requerido en los siguientes casos:
- Modulación ASK.
 - Modulación FSK. La modulación se lleva a cabo usando dos ondas senoidales de 100 y 104 Mhz.
 - Repita el inciso b) si las señales senoidales tienen una frecuencia de 100 y 120 Mhz.
- V.5. La salida de un multiplexor es de 2 400 bits/seg, la cual es la entrada de un módem. Calcule el ancho de banda de transmisión requerido en los siguientes casos:
- FSK con una desviación de frecuencia de $\pm 2\,400$ Hz alrededor de la portadora.
 - ASK.
 - QAM de 16 niveles (explique su respuesta).
 - QPSK de 8 fases (explique su respuesta).
- V.6. Un canal telefónico tiene un ancho de banda de 600 a 3 300 Hz.
- Muestre si es posible transmitir a una velocidad de 2 400 bits/seg a través de este canal, utilizando modulación PSK de 4 fases y pulsos cosenoidales.
 - Determine qué tipo de modulación y qué factor de *roll-off* se necesita para transmitir a una velocidad de 4 800 bits/seg.
- V.7. Una computadora transmite símbolos binarios a una velocidad de 56 Kbits/seg. Estos pulsos se transmiten a través de un sistema PSK conformados como pulsos senoidales. Encuentre el ancho de banda de transmisión necesario si se utiliza un factor de *roll-off* de $\alpha = 0.5$ y $\alpha = 0.25$.

BIBLIOGRAFÍA

- [1] Schwartz, M., *Information, Transmission, Modulation and Noise*, McGraw-Hill.
- [2] Bellamy, J., *Digital Telephony*, J. Wiley.
- [3] Carlson, A. B., *Communications Systems*, McGraw-Hill.
- [4] Viterbi, A., *Principles of Coherent Communication*, McGraw-Hill.
- [5] Leon W. Couch II, *Digital and Analog Communication Systems*, 4a. ed., Macmillan.

APÉNDICE A

REVISIÓN DE SEÑALES DETERMINÍSTICAS Y REDES

INTRODUCCIÓN

LAS FORMAS DE ONDA pueden clasificarse como determinísticas o aleatorias: una forma de onda determinística es aquella que puede describirse completamente por una ecuación, una gráfica o de otra manera definitiva; una señal aleatoria es aquella en la que se requiere la teoría de probabilidad para describir las características de la forma de onda; la señal por sí misma no puede describirse con precisión. En este apéndice resumiremos las características más importantes de las formas de onda determinísticas y redes lineales para procesarlas. La teoría de las señales aleatorias la revisaremos en el apéndice B.

Una forma de onda determinística también puede clasificarse como una señal de potencia o una señal de energía. Una señal de potencia tiene potencia finita cuando se promedia en el tiempo, por lo que también tiene energía infinita; una señal de energía tiene energía finita, pero potencia de cero cuando se promedia en el tiempo. Se consideran los dos tipos de señales.

1. SEÑALES DE ENERGÍA

Las señales por lo regular se describen por su *transformada de Fourier*, ya sea en el dominio del tiempo o en el dominio de la frecuencia.

Transformada de Fourier

Una señal de energía $f(t)$ y su transformada de Fourier $F(\omega)$ están relacionados por el par de transformada de Fourier

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{j\omega t} dt \quad (1.1)$$

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{-j\omega t} d\omega \quad (1.2)$$

Si $f(t)$ es una forma de onda de voltaje, $F(\omega)$ es la densidad de voltaje con unidades V/Hz y representa las amplitudes de voltaje relativas (complejas en general) de las frecuencias que constituyen $f(t)$. La ecuación (1.2) se llama la transformada inversa de Fourier. En muchas ocasiones se

usa la notación $\mathcal{F}\{\cdot\}$ y $\mathcal{F}^{-1}\{\cdot\}$ para representar la transformada directa (Fourier) y la transformada inversa, respectivamente. La doble flecha se usa también para implicar al par. Entonces

$$\mathcal{F}^{-1}\{F(\omega)\} = f(t) \leftrightarrow F(\omega) = \mathcal{F}\{f(t)\} \quad (1.3)$$

En condiciones razonables las señales de energía siempre tendrán transformada de Fourier. De hecho, la mayoría de las formas de onda con interés práctico, tanto de energía como de potencia, tendrán transformada.

Propiedades de la transformada de Fourier

En esta sección veremos las propiedades útiles de la transformada de Fourier sin pruebas. Se asume que las formas de onda $f(t)$ y $f_n(t)$ para $n = 1, 2 \dots N$ tienen transformadas de Fourier $F(\omega)$ y $F_n(\omega)$ para $n = 1, 2 \dots N$ respectivamente, mientras ω_0 y t_0 son constantes reales y α_n para $n = 1, 2 \dots N$ son posibles constantes complejas. Las formas de onda en el tiempo pueden ser complejas.

La propiedad de linealidad es:

$$f(t) = \sum_{n=1}^N \alpha_n f_n(t) \leftrightarrow \sum_{n=1}^N \alpha_n F_n(\omega) = F(\omega) \quad (1.4)$$

Las propiedades de corrimiento en el tiempo y en la frecuencia son:

$$f(t - t_0) \leftrightarrow F(\omega) e^{-j\omega t_0} \quad (1.5)$$

$$f(t) e^{j\omega_0 t} \leftrightarrow F(\omega - \omega_0) \quad (1.6)$$

Las propiedades de convolución son:

$$f(t) = \int_{-\infty}^{\infty} f_1(\tau) f_2(t - \tau) d\tau \leftrightarrow F_1(\omega) F_2(\omega) = F(\omega) \quad (1.7)$$

$$f(t) = f_1(t) f_2(t) \leftrightarrow \frac{1}{2\pi} \int_{-\infty}^{\infty} F_1(\xi) F_2(\omega - \xi) d\xi = F(\omega) \quad (1.8)$$

Entonces, la convolución de dos funciones en el dominio en el tiempo corresponde al producto de sus dos espectros en el dominio de la frecuencia. Recíprocamente, el producto de dos funciones en el dominio del tiempo corresponde a la convolución de sus dos espectros en el dominio de la frecuencia divididos por 2π .

Las propiedades de correlación son

$$f(t) = \int_{-\infty}^{\infty} f_1^*(\tau) f_2(t + \tau) d\tau \leftrightarrow F_1^*(\omega) F_2(\omega) = F(\omega) \quad (1.9)$$

$$f(t) = f_1^*(t) f_2(t) \leftrightarrow \frac{1}{2\pi} \int_{-\infty}^{\infty} F_1^*(\xi) F_2(\xi - \omega) d\xi = F(\omega) \quad (1.10)$$

donde el asterisco representa el complejo conjugado.

Otra propiedad que se llama el *teorema de Parseval* y se deriva de la ecuación (1.9) con $t = 0$:

$$\int_{-\infty}^{\infty} f_1^*(\tau) f_2(\tau) d\tau = \frac{1}{2\pi} \int_{-\infty}^{\infty} F_1^*(\omega) F_2(\omega) d\omega \quad (1.11)$$

Claramente, cuando $f_1(t) = f_2(t) = f(t)$,

$$\int_{-\infty}^{\infty} |f(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(\omega)|^2 d\omega \quad (1.12)$$

lo cual es muy importante pues muestra que la energía de $f(t)$ puede obtenerse por una integración ya sea por el dominio del tiempo o de la frecuencia.

De las ecuaciones (1.1) y (1.2) se derivan las propiedades de *área*:

$$F(0) = \int_{-\infty}^{\infty} f(t) dt \quad (1.13)$$

$$f(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) d\omega \quad (1.14)$$

Energía y espectro de densidad de energía

Si E denota la energía en $f(t)$, entonces de la expresión 1.12

$$E = \int_{-\infty}^{\infty} |f(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathcal{E}(\omega) d\omega \quad (1.15)$$

donde se define

$$\mathcal{E}(\omega) = |F(\omega)|^2 \quad (1.16)$$

como el *espectro de densidad de energía*. Claramente la unidad de $\mathcal{E}(\omega)$ es energía por hertz si $f(t)$ es una forma de onda de voltaje o corriente.

2. ALGUNAS SEÑALES DE ENERGÍA ÚTILES

Dos formas de onda de energía ocurren con mucha frecuencia, por lo que se definen e incluyen como referencia.

Función rectangular

La función rectangular (pulso) $\text{rect}(\cdot)$ se define por

$$f(t) = \text{rect}\left(\frac{t}{\tau}\right) = \begin{cases} 1, & |t| < \tau/2 \\ 0, & |t| > \tau/2 \end{cases} \longleftrightarrow \tau \text{sinc}\left(\frac{\omega\tau}{2}\right) = F(\omega) \quad (2.1)$$

también la función *sampling sinc* $\text{sinc}(\cdot)$ se define como

$$\text{sinc}(x) = \frac{\text{sen}(x)}{x} \quad (2.2)$$

La función $\text{rect}(t/\tau)$ y su espectro se ilustran en la figura A1.

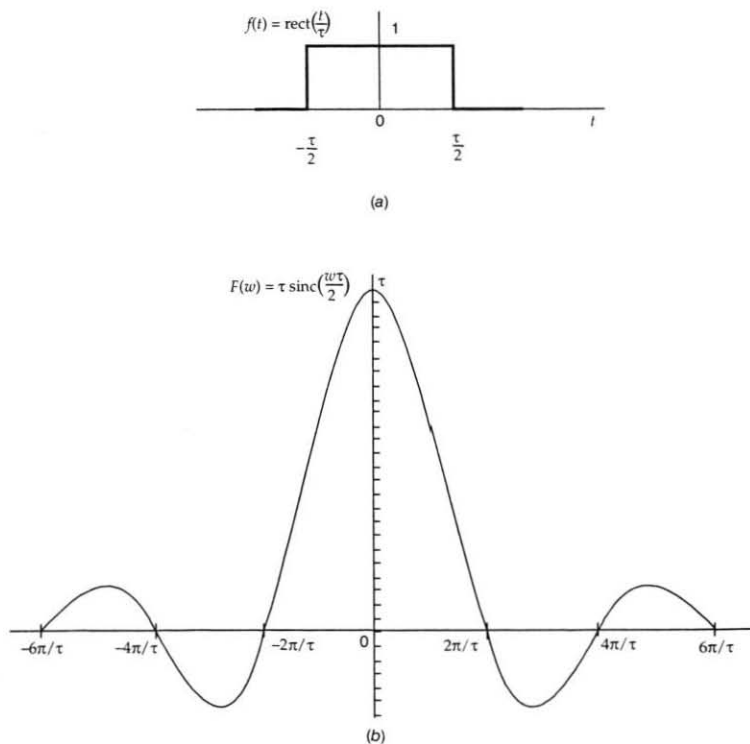


Figura A1. a) La función rectangular, b) su transformada de Fourier

Función triangular

La función triangular $\text{tri}(\cdot)$ se define por

$$f(t) = \text{tri}\left(\frac{t}{\tau}\right) = \begin{cases} 1 - |t/\tau|, & |t| \leq \tau \\ 0, & |t| > \tau \end{cases} \longleftrightarrow \tau \text{sinc}^2\left(\frac{\omega\tau}{2}\right) = F(\omega) \quad (2.3)$$

La función y su espectro se muestran en la figura A2.

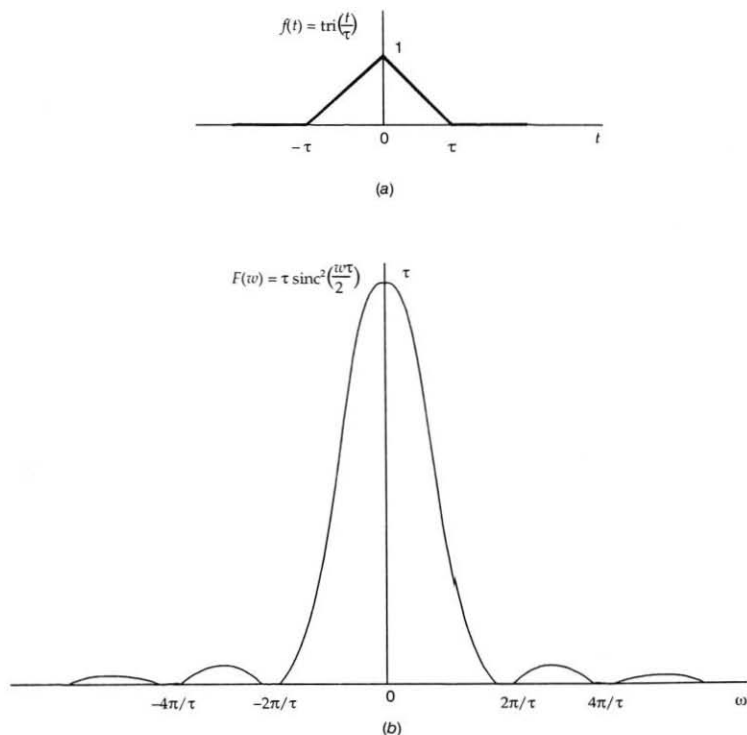


Figura A2. a) La función triangular, b) su transformada de Fourier

3. SEÑALES DE POTENCIA

Como ya dijimos, una señal de potencia tiene potencia finita cuando se promedia en el tiempo. En esta sección daremos más detalles acerca de las señales de potencia.

Potencia promedio y espectro de densidad de potencia

Se define

$$\mathcal{A}[\bullet] = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T [\bullet] dt \quad (3.1)$$

como el promedio infinito-tiempo de la cantidad entre paréntesis cuadrados. Para una forma de onda de voltaje $f(t)$, tenemos

$$\mathcal{A}[f(t)] = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f(t) dt \quad (3.2)$$

que es la componente de dc de $f(t)$, o

$$P = \mathcal{A}[|f(t)|^2] = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T |f(t)|^2 dt, \quad (3.3)$$

que es la potencia promedio P de $f(t)$. Si $f(t)$ es una forma de onda real, $|f(t)|^2$ se reemplaza por $f^2(t)$ en la ecuación (3.3).

La potencia promedio también puede encontrarse a partir de una operación en el dominio de la frecuencia. Sea $f_T(t)$ la representación de $f(t)$ truncada al intervalo $[-T, T]$ y su transformada de Fourier $F_T(\omega)$ de acuerdo con

$$f_T(t) = f(t) \operatorname{rect}\left(\frac{t}{2T}\right) \longleftrightarrow F_T(\omega) \quad (3.4)$$

Se puede demostrar que

$$P = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathcal{L}'(\omega) d\omega \quad (3.5)$$

donde

$$\mathcal{L}'(\omega) = \lim_{T \rightarrow \infty} \left\{ \frac{|F_T(\omega)|^2}{2T} \right\} \quad (3.6)$$

se conoce como el espectro de densidad de potencia de $f(t)$. La función $\mathcal{L}(\omega)$ tiene unidades de potencia por hertz.

Función de autocorrelación en el tiempo

El promedio en el tiempo del producto $f^*(t)f(t+\tau)$ definido por

$$\mathcal{R}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f^*(t) f(t+\tau) dt \quad (3.7)$$

se conoce como la función de autocorrelación en el tiempo de $f(t)$. Podemos demostrar que $\mathcal{R}(\tau)$ y $\mathcal{L}(\omega)$ están relacionadas por la transformada de Fourier, que es

$$\mathcal{L}(\omega) = \int_{-\infty}^{\infty} \mathcal{R}(\tau) e^{j\omega\tau} d\tau \quad (3.8)$$

$$\mathcal{R}(\tau) = \int_{-\infty}^{\infty} \mathcal{L}(\omega) e^{j\omega\tau} d\omega \quad (3.9)$$

entonces

$$\mathcal{R}(\tau) \leftrightarrow \mathcal{L}(\omega) \quad (3.10)$$

4. SEÑALES PERIÓDICAS DE POTENCIA

La señal periódica es uno de los tipos de señales de potencia más importantes.

Series de Fourier

Una señal periódica $f(t)$ con periodo T_p satisface la relación

$$f(t) = f(t + kT_p), \quad k = \pm 1, \pm 2, \dots \quad (4.1)$$

y puede representarse en términos de una serie infinita de funciones en el tiempo conocidas como *series de Fourier*. Las dos formas principales de las series de Fourier pueden establecerse:

Como forma trigonométrica

$$f(t) = \left(\frac{a_0}{2}\right) + \sum_{n=1}^{\infty} a_n \cos(n\omega_p t) + \sum_{n=1}^{\infty} b_n \sin(n\omega_p t) \quad (4.2)$$

donde los coeficientes de la serie de Fourier están dados por

$$a_n = \left(\frac{2}{T_p} \right) \int_{t_0 - (T_p/2)}^{t_0 + (T_p/2)} f(t) \cos(n\omega_p t) dt \quad n = 0, 1, 2 \dots \quad (4.3)$$

$$b_n = \left(\frac{2}{T_p} \right) \int_{t_0 - (T_p/2)}^{t_0 + (T_p/2)} f(t) \sin(n\omega_p t) dt \quad n = 0, 1, 2 \dots \quad (4.4)$$

con

$$\omega_p = \frac{2\pi}{T_p} \quad (4.5)$$

y t_0 es cualquier constante real.

La serie de Fourier compleja o exponencial es

$$f(t) = \sum_{n=-\infty}^{\infty} C_n e^{jn\omega_p t} \quad (4.6)$$

donde

$$C_n = \left(\frac{1}{T_p} \right) \int_{t_0 - (T_p/2)}^{t_0 + (T_p/2)} f(t) e^{-jn\omega_p t} dt \quad n = 0, \pm 1, \pm 2 \dots \quad (4.7)$$

Función impulso

Para definir y tratar con el espectro (transformada de Fourier) de una señal periódica es necesario introducir el concepto de una función impulso. La *función impulso unitario*, muchas veces llamada *función delta*, denotada por $\delta(t)$, se define por la propiedad de la integral

$$\int_{-\infty}^{\infty} \phi(t) \delta(t - t_0) dt = \phi(t_0) \quad (4.8)$$

Aquí t_0 es una constante y $\phi(t)$ es arbitraria, excepto que se asume continua en $t = t_0$. La función $\delta(t - t_0)$ se comporta como si ocurriera en $t = t_0$, tiene duración cero y amplitud infinita en $t = t_0$, y su área es unitaria. Hay otras formas impulso y funciones discontinuas $\phi(t)$ pueden manipularse.

Una función impulso $A\delta(t - t_0)$ es un impulso unitario (área uno), escalada por una amplitud constante A . Por lo regular gráficamente se muestra como una flecha vertical en el tiempo $t = t_0$ con amplitud A ; por supuesto, las puntas de flecha apuntan al infinito para implicar una amplitud infinita, mientras que la amplitud dibujada indica la constante de escala A .

Empleando la ecuación (4.8), la transformada de Fourier de la función impulso es establecida como

$$\delta(t) \leftrightarrow 1 \quad (4.9)$$

$$A \delta(t - t_0) = A e^{-j\omega t_0} \quad (4.10)$$

Otras propiedades usuales de la función impulso son, con $\omega_p = 2\pi/T_p$

$$\sum_{n=-\infty}^{\infty} \delta(t - n T_p) \leftrightarrow 2\pi \sum_{n=-\infty}^{\infty} \left(\frac{1}{T_p} \right) \delta(\omega - n \omega_p) \quad (4.11)$$

$$\sum_{n=-\infty}^{\infty} \delta(t - n T_p) = \frac{1}{T_p} \sum_{n=-\infty}^{\infty} e^{jn\omega_p t} \quad (4.12)$$

Espectro de una señal periódica

Transformando por Fourier $f(t)$ de la ecuación (4.6) se obtiene el espectro de una señal periódica

$$F(\omega) = 2\pi \sum_{n=-\infty}^{\infty} C_n \delta(\omega - n \omega_p) \quad (4.13)$$

Espectro de potencia, función autocorrelación y potencia

El espectro de potencia de una señal periódica es

$$\mathcal{P}(\omega) = 2\pi \sum_{n=-\infty}^{\infty} |C_n|^2 \delta(\omega - n \omega_p) \quad (4.14)$$

Mediante la transformada inversa de Fourier, la función de autocorrelación es

$$\mathcal{P}(\tau) = \sum_{n=-\infty}^{\infty} |C_n|^2 e^{jn\omega_p \tau} \quad (4.15)$$

La potencia promedio, resultado del uso de la ecuación (4.14) en la expresión 3.5 o de la 3.7 con $\tau = 0$:

$$P = \sum_{n=-\infty}^{\infty} |C_n|^2 \quad (4.15)$$

5. ANCHO DE LA BANDA DE LA SEÑAL Y ESPECTRO EXTENDIDO

Hay muchas definiciones acerca del ancho de banda. Generalmente, el término ancho de banda se refiere a la banda de frecuencias que son más importantes (amplitudes más grandes) para definir la señal. El espectro extendido se refiere a la banda de frecuencias fuera de la cual las componentes espectrales de la señal pueden considerarse sin importancia.

Ancho de banda de 3 dB

La banda de frecuencias positivas sobre las cuales la magnitud del espectro de la señal permanece por encima de $1/\sqrt{2}$ veces su valor a una frecuencia de referencia conveniente, usualmente localizada en la banda, se llama ancho de banda de 3 dB. La figura A3 ilustra el ancho de banda de 3 dB indicado como W en radianes por segundo para varias formas de espectros. La frecuencia de referencia para una señal en banda base (pasa bajas), mostrada en *a*, usualmente es cero, aunque también puede ser otro valor conveniente (por ejemplo, 1000 Hz en la banda de audio). Algunos ejemplos de señales pasa banda se muestran en *b* y *c*. El espectro extendido, indicado como W_f , lo

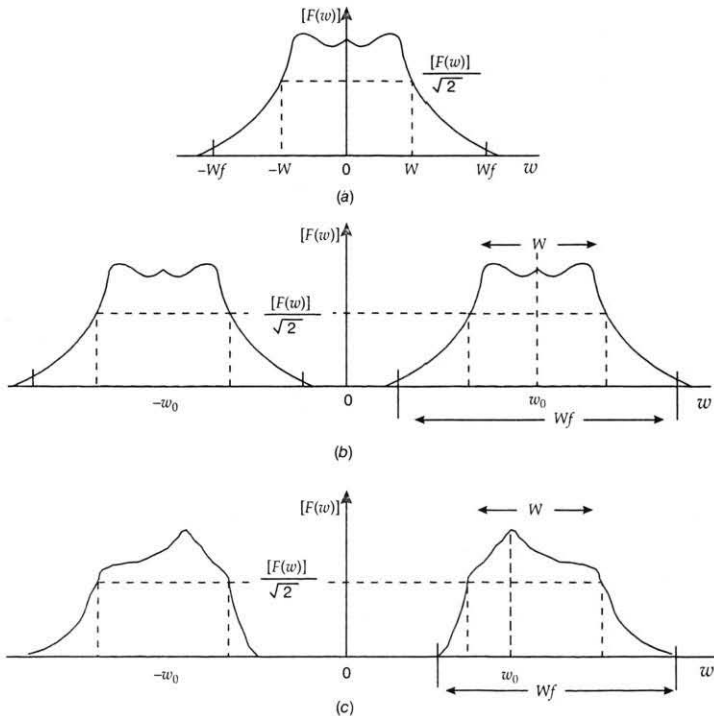


Figura A3. Espectro de señales: a) Pasa bajas o banda base, b) pasa bandas, y c) pasa bandas absolutamente limitada

vemos en a y b para señales típicas pasa bandas y pasa bajas. Una forma de onda cuyo espectro es absolutamente cero fuera de una banda como se muestra en c , se dice que está totalmente limitada en banda, y su ancho de banda absoluto, en este caso coincide con el espectro extendido.

En el caso donde $f(t)$ es periódica, la magnitud del espectro consta de impulsos. El cálculo del ancho de banda es el mismo excepto que se usa la envolvente de la magnitud del espectro.

Frecuencia media y ancho de banda RMS

Sea $F(\omega)$ el espectro de una señal $f(t)$ (pasa bandas o pasa bajas) la frecuencia media ω_0 se define por

$$\bar{\omega}_0 = \frac{\int_0^{\infty} \omega |F(\omega)|^2 d\omega}{\int_0^{\infty} |F(\omega)|^2 d\omega} \quad (5.1)$$

y el ancho de banda RMS W_{rms} por

$$W_{\text{rms}}^2 = \frac{\int_0^{\infty} (\omega - \bar{\omega}_0)^2 |F(\omega)|^2 d\omega}{\int_0^{\infty} |F(\omega)|^2 d\omega} \quad (5.2)$$

para señales pasa banda. Para formas de onda pasa bajas (ecuación (5.2)) se aplica si $\bar{\omega}_0$ es cero.

6. REDES LINEALES

Respuesta al impulso

Una red que tenga un puerto de salida y responda a una señal como excitación en un puerto de entrada, es lineal si se aplica la superposición. Un impulso unitario $\delta(t - \tau)$ aplicada a una red lineal causa una respuesta que es característica de la red, conocida como la respuesta al impulso $h(t, \tau)$. La respuesta de la red $g(t)$ a una entrada arbitraria $f(t)$ está dada por

$$g(t) = \int_{-\infty}^{\infty} f(\tau) h(t, \tau) d\tau \quad (6.1)$$

Si la red también es invariante en el tiempo, de forma tal que $h(t, \tau)$ no depende en lo absoluto de la forma en que ocurre la entrada del impulso en el tiempo, entonces

$$h(t, \tau) = h(t - \tau) \quad (6.2)$$

y

$$g(t) = \int_{-\infty}^{\infty} f(\tau) h(t - \tau) d\tau \quad (6.3)$$

Esta expresión se aplica en general a señales de potencia y de energía.

Función de transferencia

Mediante la transformada de Fourier de la ecuación (6.3), tenemos

$$G(\omega) = F(\omega) H(\omega) \quad (6.4)$$

donde $G(\omega)$, $F(\omega)$, y $H(\omega)$ son respectivamente las transformadas de $g(t)$, $f(t)$ y $h(t)$. $H(\omega)$ es la función de transferencia de la red invariante en el tiempo. La ecuación (6.4) se aplica a todas las señales (potencia, energía), para las que puede definirse $F(\omega)$.

Ancho de banda

El ancho de banda de una red (3 dB o *rms*) puede definirse de la misma forma que para formas de onda. Sólo es necesario utilizar $|H(\omega)|$ en lugar de $|F(\omega)|$, que usamos en la sección anterior.

Redes ideales

Frecuentemente se desea aproximar la función de transferencia de un filtro real por un *filtro ideal*. Un filtro ideal se define como aquel que tiene una ganancia unitaria y función de transferencia perfectamente plana (constante) en el intervalo deseado y respuesta cero en cualquiera de las otras frecuencias. Esta definición puede entenderse usando la figura 6.1, que ilustra filtros pasa bajas, pasa altas y pasa bandas ideales. En cada caso el filtro puede tener una fase lineal $\theta(\omega) = -\omega\tau$, donde τ es la constante de retardo del filtro.

Respuesta de espectro de potencia y energía

Si la entrada $f(t)$ a una red lineal invariante en el tiempo es una señal de energía con espectro de densidad de energía $E_f(\omega) = |F(\omega)|^2$, la densidad de energía de la respuesta $g(t)$, denotada por $E_g(\omega)$, es

$$E_g(\omega) = |G(\omega)|^2 = E_f(\omega) |H(\omega)|^2 \quad (6.5)$$

de la ecuación (6.4). La energía total E_g de la respuesta es

$$E_g = \frac{1}{2\pi} \int_{-\infty}^{\infty} E_f(\omega) |H(\omega)|^2 d\omega \quad (6.6)$$

Si $f(t)$ es una señal de potencia, con espectro de densidad de potencia $L_f(\omega)$, el espectro de potencia $L_g(\omega)$, de $g(t)$ está dado por

Apéndice A

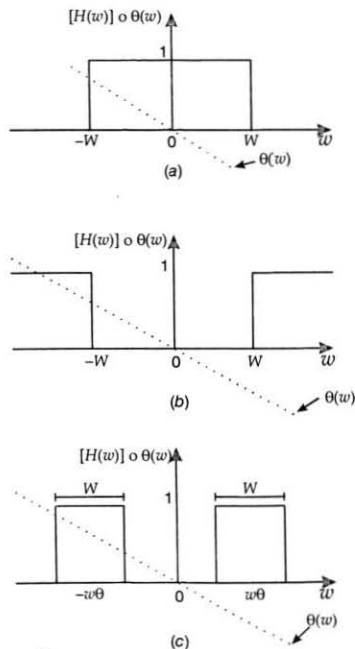


Figura A4. Filtros. a) Pasa bajos, b) pasa altos y c) pasa bandas ideales

$$L_g(\omega) = L_f(\omega) |H(\omega)|^2 \quad (6.7)$$

La potencia de salida P_g en $g(t)$ es

$$P_g = \frac{1}{2\pi} \int_{-\infty}^{\infty} L_f(\omega) |H(\omega)|^2 d\omega \quad (6.8)$$

APÉNDICE B

REVISIÓN DE TEORÍA DE SEÑALES ALEATORIAS

INTRODUCCIÓN

EN ESTE APÉNDICE haremos una pequeña revisión de los elementos básicos de la teoría de señales aleatorias, necesarias para la comprensión del texto. Las palabras *señal aleatoria* se emplean para designar ya sea una forma de onda aleatoria deseada (referida usualmente como la señal) o una forma de onda aleatoria no deseada, como el ruido.

La revisión comienza a partir del concepto fundamental de probabilidad, a partir del cual se basa todo lo demás.

1. ESPACIOS MUESTRALES, EVENTOS Y PROBABILIDAD

Espacios muestrales

Un experimento para el que los resultados de las pruebas son aleatorios se llama un *experimento aleatorio*. Un ejemplo de experimento aleatorio es tomar una carta de un mazo de 52 cartas. Aquí la prueba es tomar una carta y la carta es el resultado; en este caso hay 52 posibles resultados para este experimento. El conjunto de todos los posibles resultados en un experimento aleatorio es el espacio muestral y los resultados son los elementos del espacio muestral. En el experimento de la carta el espacio muestral es *discreto*, ya que sólo son posibles resultados discretos.

Experimentos diferentes pueden dar diferentes espacios muestrales. Considere un potenciómetro circular cuyo tap tenga un giro de 360° y que genere cualquier voltaje desde 0 (posición de cero grados) hasta, por ejemplo, 15 V (posición de 360°). Un experimento que consiste en ubicar aleatoriamente el tap produce un resultado (voltaje) que puede tener cualquier valor en un continuo de valores, desde 0 hasta 15 V. El espacio muestral de este experimento es llamado continuo.

Eventos

Un evento es definido como un subconjunto del espacio muestral. En el experimento de las cartas podemos estar más interesados en obtener un rey que cualquier otra carta. Aquí cuatro elementos del espacio muestral satisfacen el evento (ya que hay 4 reyes en el mazo).

Probabilidad

Probabilidad es una función de los eventos definidos en el espacio muestral. Se trata de un número asignado a cada evento que define la posibilidad relativa de que éste pueda ocurrir. La asignación se hace con sentido común, del resultado de mediciones o de cualquier otra cosa razonable y justificable; por ejemplo es razonable que en el experimento de las cartas el evento *seleccionar un rey* tenga la probabilidad $4/52$, ya que hay 4 reyes en un total de 52 cartas, y se presume que todas pueden salir igualmente.

Si los eventos se denotan por letras mayúsculas (y los elementos de los eventos en minúsculas) como A, B, C , la probabilidad del evento será $P(A)$.

Probabilidad conjunta

Se denota la probabilidad de que dos eventos, A y B , puedan ocurrir al mismo tiempo o simultáneamente en una prueba de un experimento, por $P(A, B)$. Por ejemplo, si en una caja hay 90 resistencias mezcladas, como se muestra en el cuadro 1, y si A es el evento “tomar una resistencia de $100\ \Omega$ ” y B es el evento “tomar una resistencia de $2\ W$ ”, entonces $P(A, B) = 9/90 = 0.1$. Es importante observar que $P(A) = 32/90$ y $P(B) = 20/90$ para este ejemplo.

CUADRO 1. Resistencias en una caja ordenadas por resistencia y potencia

Resistencia	Potencia			Totales
	1/2 W	1 W	2 W	
10 Ω	8	6	5	19
100 Ω	10	13	9	32
1 000 Ω	18	15	6	39
Totales	36	34	20	90

Para varios eventos $A_1, A_2 \dots A_N$, su probabilidad conjunta se denota por $P(A_1, A_2 \dots A_N)$.

Probabilidad condicional

Algunas veces estamos interesados en que la probabilidad de un evento, como A , ocurra, puesto que otro evento, como B , ya ocurrió. A esto se le conoce como probabilidad condicional y se denota por $P(A/B)$. En el ejemplo del cuadro 1 tenemos $P(A/B) = 9/20$. La probabilidad conjunta y la condicional están relacionadas por

$$P(A, B) = P(A|B) P(B) \quad (1.1)$$

o alternativamente

$$P(A, B) = P(B|A) P(A) \quad (1.2)$$

Combinando las ecuaciones (1.2) y (1.1) obtenemos una forma de la regla de Bayes

$$P(A, B) = P(B|A)P(A) \quad (1.3)$$

Independencia estadística

Dos eventos, A y B , se dice que son estadísticamente independientes si

$$P(A|B) = P(A), \text{ eventos independientes} \quad (1.4)$$

Una definición equivalente es

$$P(A, B) = P(A)P(B), \text{ eventos independientes} \quad (1.5)$$

En el experimento anterior, “tomar una resistencia de la caja” se tiene $P(A) = 32/90$ y $P(A|B) = 9/20$, entonces $P(A|B) \neq P(A)$, lo que significa que los eventos A y B no son independientes.

Múltiples eventos $A_1, A_2 \dots A_N$ se dice que son estadísticamente independientes si todas las condiciones

$$\begin{aligned} P(A_i, A_j) &= P(A_i)P(A_j) \\ P(A_i, A_j, A_k) &= P(A_i)P(A_j)P(A_k) \\ &\vdots \end{aligned} \quad (1.6)$$

$$P(A_i, A_j \dots A_N) = P(A_i)P(A_j) \dots P(A_N)$$

son satisfechas para $1 \leq i < j < k < \dots \leq N$. Aquí hay $2^N - N - 1$ de estas condiciones.

2. VARIABLES ALEATORIAS, DISTRIBUCIONES Y DENSIDADES

En algunos experimentos aleatorios los resultados no son numéricos, tal como en los experimentos de las resistencias y cartas. El concepto *variable aleatoria* permite que experimentos aleatorios se representen numéricamente.

Variable aleatoria

Una variable aleatoria se define como una función real de los elementos de un espacio muestral; estos puntos (elementos) de un espacio muestral se mapearan como puntos (números) en la recta

real de la función. Las letras mayúsculas (W, X o Y) se usan para representar variables aleatorias, y un valor particular de la variable aleatoria se denota por letras minúsculas (w, x o y).

Las variables aleatorias pueden ser discretas, continuas, o una mezcla de las dos, dependiendo del espacio muestral y de la forma de la función que define la variable aleatoria.

Funciones de distribución

Sea A un conjunto de puntos en el espacio muestral que se mapean en un conjunto de puntos en la recta real $|X \leq x|$. Esos puntos corresponden a todos los valores de la variable aleatoria X que no excedan un número arbitrario x . La probabilidad del conjunto A definido en el espacio muestral es igual a la probabilidad del conjunto $|X \leq x|$, debido al mapeo punto a punto. Esta probabilidad es una función de x , que se denota por $P_x(x)$, y está dada por

$$P_x(x) = P |X \leq x| \quad (2.1)$$

La función $P_x(x)$ es llamada la función de distribución de probabilidad acumulativa de X . Ésta es una función que tiene las siguientes propiedades:

$$P_x(-\infty) = 0 \quad (2.2a)$$

$$P_x(\infty) = 1.0 \quad (2.2b)$$

$$0 \leq P_x(x) \leq 1.0 \quad (2.2c)$$

$$P_x(x_1) \leq P_x(x_2) \text{ si } x_1 < x_2 \quad (2.2d)$$

$$P |x_1 < X \leq x_2| = P_x(x_2) - P_x(x_1) \quad (2.2e)$$

Para una variable aleatoria discreta, $P_x(x)$ tendrá escalones con amplitudes iguales a las probabilidades de los posibles valores discretos de X , denotados por x_i , $i = 1, 2 \dots N$:

$$P_x(x) = \sum_{i=1}^N P |X = x_i| u(x - x_i) \quad (2.3)$$

la función escalón unitario se define por

$$u(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (2.4)$$

Cuando se involucra más de una variable aleatoria, la función de distribución de probabilidad conjunta está definida por

$$P_{x_1, x_2, \dots, x_N}(x_1, x_2, \dots, x_N) = P |X_1 \leq x_1, X_2 \leq x_2, \dots, X_N \leq x_N| \quad (2.5)$$

Funciones de densidad de probabilidad

La función de densidad de probabilidad, denotada por $p_x(x)$ se define como la derivada de la distribución

$$p_x(x) = \frac{d P_x(x)}{dx} \quad (2.6)$$

las propiedades de $p_x(x)$ son

$$0 \leq p_x(x) \quad (2.7a)$$

$$\int_{-\infty}^{\infty} p_x(x) dx = 1 \quad (2.7b)$$

$$P_x(x) = \int_{-\infty}^x p_x(\xi) d\xi = 1 \quad (2.7c)$$

$$P_x(x_1 < X \leq x_2) = \int_{x_1}^{x_2} p_x(x) dx = 1 \quad (2.7d)$$

Para una variable aleatoria discreta,

$$p_x(x) = \sum_{i=1}^N P[X = x_i] \delta(x - x_i) \quad (2.8)$$

Para varias variables aleatorias, la función de densidad de probabilidad conjunta es

$$p_{x_1, x_2, \dots, x_N}(x_1, x_2, \dots, x_N) = \frac{\partial^N P_{x_1, x_2, \dots, x_N}(x_1, x_2, \dots, x_N)}{\partial x_1 \partial x_2 \dots \partial x_N} \quad (2.9)$$

Densidad y distribución condicional

Nuevamente sea el evento $A = \{X \leq x\}$, con la ecuación (1.1) tenemos la probabilidad condicional de $\{X \leq x\}$ de un evento B que haya ocurrido. A esta probabilidad se le llama distribución condicional de X , denotada por $P_x(x/B)$:

$$P_x(x/B) = P[X \leq x/B] = \frac{P[X \leq x, B]}{P(B)} \quad (2.10)$$

La densidad condicional sigue la derivada

$$p_x(x|B) = \frac{dP_x(x|B)}{dx} \quad (2.11)$$

Cuando varias variables aleatorias, $X_1, X_2 \dots X_N$ son consideradas, se extienden las ecuaciones (2.10) y (2.11) para obtener

$$P_{x_1, x_2 \dots x_N}(x_1, x_2 \dots x_N|B) = \frac{P[X_1 \leq x_1, X_2 \leq x_2 \dots X_N \leq x_N, B]}{P(B)} \quad (2.12)$$

$$p_{x_1, x_2 \dots x_N}(x_1, x_2 \dots x_N|B) = \frac{\partial^N P_{x_1, x_2 \dots x_N}(x_1, x_2 \dots x_N|B)}{\partial x_1 \partial x_2 \dots \partial x_N} \quad (2.13)$$

En el caso de que el evento B en la expresión 2.10 se defina por una segunda variable aleatoria Y de acuerdo con $B = [y - \Delta y < Y \leq y + \Delta y]$ con $\Delta y \rightarrow 0$, de forma tal que Y se acerca a un valor fijo y , se puede mostrar que

$$P_x(x|y) = \lim_{\Delta y \rightarrow 0} P[X \leq x | y - \Delta y < Y \leq y + \Delta y] = \frac{\int_{-\infty}^{\infty} p_{X,Y}(\xi, y) d\xi}{p_Y(y)} \quad (2.14)$$

$$p_x(x|y) = \frac{p_{X,Y}(x, y)}{p_Y(y)} \quad (2.15)$$

Independencia estadística

N variables aleatorias $X_i, i = 1, 2 \dots N$ se dice que son estadísticamente independientes si sus distribuciones conjuntas (o densidades) satisfacen las condiciones de la ecuación (1.6). Una condición particular

$$p_{x_1, x_2 \dots x_N}(x_1, x_2 \dots x_N) = \prod_{i=1}^N p_{x_i}(x_i) \quad (2.16)$$

es sumamente útil en el análisis de muchos sistemas prácticos.

3. PROMEDIOS ESTADÍSTICOS

Promedio de una función de variables aleatorias

Sea $g(X_1, X_2 \dots X_N)$ una función real de las N variables aleatorias $X_i, i = 1, 2 \dots N$. Se define el *promedio estadístico* (también llamado valor medio o valor esperado) de $g(\cdot, \dots, \cdot)$, denotado ya sea por \bar{g} o $E[g(\cdot, \dots, \cdot)]$ por

Apéndice B

$$\bar{g} = E [g (X_1, X_2 \dots X_N)] \quad (3.1)$$

$$= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g (x_1, x_2 \dots x_N) p_{x_1, x_2 \dots x_N} (x_1, x_2 \dots x_N) dx_1 dx_2 \dots dx_N$$

La mayoría de los problemas prácticos requieren solamente de dos variables aleatorias, digamos X y Y . En este caso especial

$$\bar{g} = E [g (X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g (x, y) p_{X, Y} (x, y) dx dy \quad (3.2)$$

Momentos

Los promedios estadísticos de un número especial de funciones $g (X, Y)$ son especialmente importantes. Estos promedios se llaman momentos de las variables aleatorias X y Y ; por ejemplo,

$$\bar{X} = E [X] = \int_{-\infty}^{\infty} x p_X (x) dx \quad (3.3)$$

llamada la media o primer momento de X (en el origen) y

$$\overline{X^2} = E [X^2] = \int_{-\infty}^{\infty} x^2 p_X (x) dx \quad (3.4)$$

es la potencia en X (segundo momento en el origen). Otro momento (alrededor de la media de X) es la varianza de X y se denota por σ_X^2 , dada por

$$\sigma_X^2 = E [(X - \bar{X})^2] = \int_{-\infty}^{\infty} (x - \bar{X})^2 p_X (x) dx \quad (3.5)$$

Otros momentos importantes son la correlación y la covarianza de X y Y , denotados por R_{XY} y C_{XY} respectivamente. Estas cantidades están definidas por

$$R_{XY} = \overline{XY} = E [XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy p_{X, Y} (x, y) dx dy \quad (3.6)$$

$$C_{XY} = \overline{(X - \bar{X})(Y - \bar{Y})} = E [(X - \bar{X})(Y - \bar{Y})] \quad (3.7)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{X})(y - \bar{Y}) p_{X, Y} (x, y) dx dy$$

Las siguientes relaciones son de suma importancia

$$\sigma_X^2 = \overline{X^2} - (\overline{X})^2 \quad (3.8)$$

$$C_{XY} = \overline{XY} - \overline{X} \overline{Y} = R_{XY} - \overline{X} \overline{Y} \quad (3.9)$$

4. VARIABLES ALEATORIAS GAUSSIANAS

Una variable aleatoria X es gaussiana si su función de densidad de probabilidad está dada por

$$p_X(x) = (2\pi\sigma_X^2)^{-1/2} \exp\left[-\frac{(x - \overline{X})^2}{2\sigma_X^2}\right] \quad (4.1)$$

Esta función la vemos graficada en la figura B.1; es una función simétrica alrededor de la media de X y se expande proporcionalmente a σ_X (llamada desviación estándar de X). La densidad gaussiana normalizada resulta cuando $\overline{X} = 0$ y $\sigma_X^2 = 1$. Empleando la notación $p(x)$ para el caso normalizado, la función de distribución normalizada es

$$P(x) = \int_{-\infty}^x p(x) dx \quad (4.2)$$

con

$$p(x) = (2\pi)^{-1/2} \exp\left[-\frac{x^2}{2}\right] \quad (4.3)$$

Se puede demostrar que la función de distribución de la variable aleatoria gaussiana está relacionada con $P_X(x)$ por

$$P_X(x) = \int_{-\infty}^x p_X(\xi) d\xi = P\left(\frac{x - \overline{X}}{\sigma_X}\right) \quad (4.4)$$

La función error, función error complementaria y función Q se definen a partir de la densidad gaussiana normalizada por

$$\operatorname{erf}(\beta) = \frac{2}{\sqrt{\pi}} \int_0^\beta e^{-\xi^2} d\xi, \quad \operatorname{erf}(-\beta) = -\operatorname{erf}(\beta) \quad (4.5)$$

$$\operatorname{erfc}(\beta) = 1 - \operatorname{erf}(\beta), \quad \operatorname{erfc}(-\beta) = 2 - \operatorname{erfc}(\beta) \quad (4.6)$$

$$Q(\beta) = \frac{1}{\sqrt{2\pi}} \int_\beta^\infty e^{-\xi^2/2} d\xi, \quad Q(-\beta) = 1 - Q(\beta) \quad (4.7)$$

Apéndice B

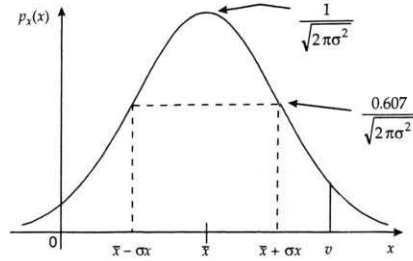


Figura B1.

Estas funciones están relacionadas por la función de distribución

$$\begin{aligned}
 P_x(x) &= 1 - Q\left(\frac{x - \bar{X}}{\sigma_x}\right) = 1 - \frac{1}{2} \operatorname{erfc}\left(\sqrt{\frac{X}{2\sigma_x}}\right) \\
 &= \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{x - \bar{X}}{\sqrt{2}\sigma_x}\right)
 \end{aligned}
 \tag{4.8}$$

5. PROCESOS Y SEÑALES ALEATORIAS

El propósito de toda la teoría que hemos dado es proporcionar las bases para describir las formas de onda aleatorias. Para usar efectivamente las bases, es preciso introducir el concepto de modelado de una señal aleatoria real por un proceso aleatorio real.

Concepto de proceso aleatorio

Considere una forma de onda aleatoria real (ruido), tal como la que se muestra en la figura B.2 como $x_0(t)$, que puede existir en algún punto dado de un sistema. Otro sistema construido idénticamente al primero, puede generar una función aleatoria diferente en el mismo punto, mostrada como $x_1(t)$ también en esta figura B.2. De la misma forma otros sistemas construidos idénticamente, podrían generar otras formas de onda, como $x_2(t)$, $x_3(t)$, y así sucesivamente. En un principio se puede imaginar una infinidad de sistemas idénticos, cada uno generando sus propias formas de onda. El conjunto de estas formas de onda se llama *ensamble*; las formas de onda son diferentes con el tiempo ya que la agitación térmica de los electrones en los conductores y semiconductores en los sistemas, que ocasiona el ruido, es diferente en cada sistema, aunque los sistemas se hayan construido de manera idéntica.

Otra manera de ver las formas de onda de la figura B.2, especialmente atractiva, es imaginarlas como las posibles formas de onda que pueden generarse por una red. El ensamble de todas las posibles formas de onda se conoce como proceso aleatorio. La señal aleatoria real solamente es un miembro del ensamble de los procesos aleatorios de las posibles señales que, tomadas juntas, describen las propiedades estadísticas de la señal; por ejemplo, al tiempo t_1 el voltaje de cualquier forma de onda se trata como un valor específico de la variable aleatoria, describiendo todos los

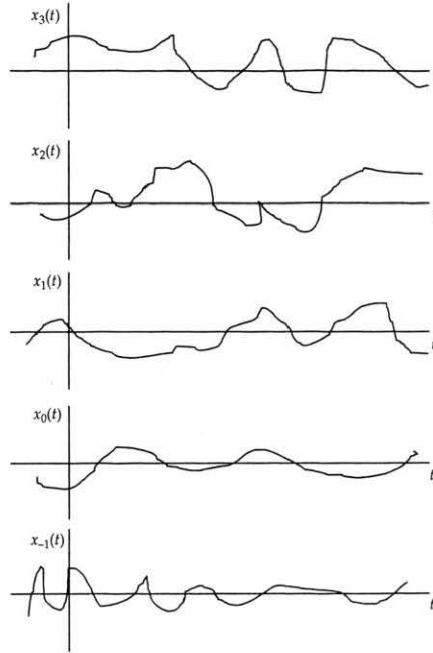


Figura B2. Un ensemble de formas de onda aleatorias $x_i(t)$, $i = 0, \pm 1, \pm 2, \dots$ que comprende un proceso aleatorio

valores de las formas de onda de los miembros del proceso. El valor medio de esta variable aleatoria es un promedio estadístico de todos los posibles voltajes que ocurren al tiempo t_1 ; esto se conoce como el *promedio del ensemble* al tiempo t_1 .

Si el proceso aleatorio se denota por $X(t)$, entonces el proceso se interpreta como una variable aleatoria X al tiempo t . La densidad de probabilidad se denota por $p_X(x; t)$. La amplitud promedio de la señal aleatoria (valor de *dc*) al tiempo t se convierte en

$$E[X(t)] = \int_{-\infty}^{\infty} x p_X(x; t) dx \quad (5.1)$$

Si $p_X(x; t)$ varía con el tiempo, entonces $E[X(t)]$ podría variar con el tiempo. Usando la misma lógica se define la potencia en el proceso por

$$E[X^2(t)] = \int_{-\infty}^{\infty} x^2 p_X(x; t) dx \quad (5.2)$$

que, en general, varía con el tiempo.

Funciones correlacionadas

Si las variables aleatorias definidas por el proceso $X(t)$ a los tiempos t_1 y t_2 son $X_1 = X(t_1)$ y $X_2 = X(t_2)$ con densidad de probabilidad conjunta denotada por $p_X(x_1, x_2; t_1, t_2)$, la correlación entre X_1 y X_2 denotada por $R_{XX}(t_1, t_2)$, es

$$R_{XX}(t_1, t_2) = E[X(t_1)X(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 p_X(x_1, x_2; t_1, t_2) dx_1 dx_2 \quad (5.3)$$

Ya que la correlación es de dos variables aleatorias tomadas del mismo proceso, ésta se conoce como la función de autocorrelación. Con $t_1 = t$ y $t_2 = t + \tau$, donde $\tau = t_2 - t_1$ es la diferencia de tiempos, tenemos

$$R_{XX}(t, t + \tau) = E[X(t)X(t + \tau)] \quad (5.4)$$

Si dos procesos aleatorios $X(t)$ y $Y(t)$ son considerados, la correlación entre las variables aleatorias $X = X(t)$ y $Y = Y(t + \tau)$ es llamada la *función de correlación cruzada* del proceso. Y está dada por

$$R_{XY}(t, t + \tau) = E[X(t)Y(t + \tau)] \quad (5.5)$$

Alternativamente

$$R_{YX}(t, t + \tau) = E[Y(t)X(t + \tau)] \quad (5.6)$$

Estacionariedad

Las funciones media, potencia y autocorrelación son todas las medidas de las propiedades estadísticas del proceso, aunque también hay otras posibles mediciones (momentos de mayor orden). Hablando en un sentido más amplio, si las propiedades estadísticas del proceso no varían con el tiempo éste se llama estacionario. Aunque hay muchas definiciones precisas de estacionariedad, una de ellas es la más útil, que se conoce como estacionariedad en el sentido amplio. Un proceso aleatorio $X(t)$ es estacionario en el sentido amplio si las siguientes dos condiciones son verdaderas:

$$E[X(t)] = \bar{X} = \text{constante} \quad (5.7a)$$

$$E[X(t)X(t + \tau)] = R_{XX}(\tau) \quad (\text{independiente de } t) \quad (5.7b)$$

Dos procesos son conjuntamente estacionarios en el sentido amplio si cada uno es estacionario en el sentido amplio y sus funciones de correlación cruzadas son absolutamente independientes del tiempo

$$E[X(t)Y(t + \tau)] = R_{XY}(\tau) \quad (5.8)$$

$$E [Y(t)X(t + \tau)] = R_{YX}(\tau) \quad (5.9)$$

Un proceso ergódico es una forma más restringida de un proceso aleatorio, y es aquel cuyos promedios en el tiempo de un miembro simple del ensamble de los procesos es igual a sus promedios estadísticos correspondientes. Si los promedios en el tiempo están definidos por

$$\mathcal{A}[\bullet] = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T [\bullet] dt \quad (5.10)$$

como usualmente se hace, entonces para el proceso ergódico

$$\bar{X} = E[X(t)] = \mathcal{A}[x(t)] = \bar{x} \quad (5.11)$$

$$\overline{X^2} = E[X^2(t)] = \mathcal{A}[x^2(t)] = \bar{x}^2 \quad (5.12)$$

$$R_{XX}(\tau) = E[X(t)X(t + \tau)] = \mathcal{A}[x(t)x(t + \tau)] = \mathcal{R}_{XX}(\tau) \quad (5.13)$$

donde se usa \mathcal{A} para representar la función de correlación en el tiempo.

Para procesos ergódicos conjuntos

$$R_{XY}(\tau) = \mathcal{R}_{xy}(\tau) \triangleq \mathcal{A}[x(t)y(t + \tau)] \quad (5.14)$$

$$R_{YX}(\tau) = \mathcal{R}_{yx}(\tau) \triangleq \mathcal{A}[y(t)x(t + \tau)] \quad (5.15)$$

Aunque los modelos estadísticos por lo regular se emplean en modelado y análisis teóricos de sistemas, también son los promedios en el tiempo más comúnmente medidos (aproximadamente en la práctica). Si el proceso es ergódico, los dos conjuntos de resultados son iguales y los resultados medidos coincidirán con la teoría. En la práctica, los procesos por lo tanto se asumen como ergódicos sin realmente probar lo que se asume; sin embargo, si el proceso no es ergódico, la validez de lo que se asume es de pequeñas consecuencias si los resultados teóricos y prácticos caen en rangos aceptables; usualmente así sucede.

6. ESPECTRO DE DENSIDAD DE POTENCIA

En la sección anterior comentamos las formas en las que se describen los procesos aleatorios usando el dominio en el tiempo. Como con señales determinísticas, las señales aleatorias se pueden describir en el dominio de la frecuencia. Esta descripción involucra el espectro de densidad de potencia en lugar del espectro de densidad de voltaje.

Procesos estacionarios

Si $X_T(t)$ representa un proceso aleatorio $X(t)$ truncado en $-T < t < T$ (y cero para $|t| > T$), entonces se puede demostrar que el proceso tiene un *espectro de densidad de potencia*

$$\mathcal{L}_{XX}(\omega) = \lim_{T \rightarrow \infty} \frac{E[|X_T(\omega)|^2]}{2T} \quad (6.1)$$

donde $X_T(\omega)$ es la transformada de Fourier de $X_T(t)$ y se asume que el proceso es al menos estacionario en el sentido amplio. El espectro de potencia de ω para procesos reales es una función real, no negativa, par.

Para procesos conjuntamente estacionarios, $X(t)$ y $Y(t)$, teniendo representaciones truncadas $X_T(t)$ y $Y_T(t)$ con sus respectivas transformadas de Fourier $X_T(\omega)$ y $Y_T(\omega)$, los espectros de densidad de potencia cruzados se pueden derivar de la siguiente manera

$$\mathcal{L}_{XY}(\omega) = \lim_{T \rightarrow \infty} \frac{E[X_T^*(\omega) Y_T(\omega)]}{2T} \quad (6.2)$$

$$\mathcal{L}_{YX}(\omega) = \lim_{T \rightarrow \infty} \frac{E[Y_T^*(\omega) X_T(\omega)]}{2T} \quad (6.3)$$

La potencia y el espectro de densidad de potencia cruzado para procesos estacionarios conjuntos en el sentido amplio están relacionados por las funciones de correlación por pares de transformadas de Fourier

$$R_{XX}(\tau) \leftrightarrow \mathcal{L}_{XX}(\omega) \quad (6.4)$$

$$R_{YY}(\tau) \leftrightarrow \mathcal{L}_{YY}(\omega) \quad (6.5)$$

$$R_{XY}(\tau) \leftrightarrow \mathcal{L}_{XY}(\omega) \quad (6.6)$$

$$R_{YX}(\tau) \leftrightarrow \mathcal{L}_{YX}(\omega) \quad (6.7)$$

Procesos no estacionarios

Aunque un proceso aleatorio no sea estacionario, el espectro de densidad de potencia promedio puede definirse como la transformada de Fourier del promedio en el tiempo; la función del proceso de autocorrelación es

$$\mathcal{L}_{XX}(\omega) = \int_{-\infty}^{\infty} \mathcal{A}[R_{XX}(t, t + \tau)] e^{-j\omega\tau} d\tau \quad (6.8)$$

De este modo

$$\mathcal{A}[R_{XX}(t, t + \tau)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathcal{L}_{XX}(\omega) e^{j\omega\tau} d\omega \quad (6.9)$$

En el caso de no estacionarios la función de autocorrelación no puede recuperarse de su espectro de potencia promedio; sólo puede recuperarse la función de promedio en el tiempo autocorrelacionada.

De la misma manera, los espectros de densidad promedio de potencia cruzada pueden definirse como las transformadas de Fourier de la función de promedio en el tiempo de correlación cruzada.

Potencia

La potencia en un proceso aleatorio $X(t)$ se deriva ya sea de su función de autocorrelación o de su espectro de densidad de potencia:

$$\begin{aligned}\overline{X^2} &= E[X^2(t)] = R_{XX}(0) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathcal{L}_{XX}(\omega) d\omega\end{aligned}\quad (6.10)$$

7. RESPUESTA DE LAS REDES A SEÑALES ALEATORIAS

Resultado fundamental

Una señal aleatoria $x(t)$ aplicada a la entrada de una red lineal invariante en el tiempo, con una sola entrada y una sola salida, generará una salida o respuesta dada por la convolución de $x(t)$ con la respuesta al impulso de la red $h(t)$:

$$\begin{aligned}y(t) &= \int_{-\infty}^{\infty} h(\xi) x(t - \xi) d\xi \\ &= \int_{-\infty}^{\infty} x(\xi) h(t - \xi) d\xi\end{aligned}\quad (7.1)$$

El resultado fundamental de la ecuación (7.1) es todo lo que se requiere para derivar todas las características de la respuesta aleatoria $y(t)$.

Ya que la forma de onda de un proceso aleatorio $X(t)$ se afecta por la red, mediante la ecuación (7.1) las respuestas corresponderán a los miembros de un nuevo proceso aleatorio $Y(t)$. Como consecuencia de este hecho se puede interpretar esta misma ecuación como la transformación de un proceso en otro

$$\begin{aligned}Y(t) &= \int_{-\infty}^{\infty} h(\xi) X(t - \xi) d\xi \\ &= \int_{-\infty}^{\infty} X(\xi) h(t - \xi) d\xi\end{aligned}\quad (7.2)$$

Funciones de correlación de salida

Si $X(t)$ es estacionario en el sentido amplio, y la red es lineal e invariante en el tiempo, se puede demostrar que el proceso de respuesta $Y(t)$ es estacionario en el sentido amplio y $X(t)$ y $Y(t)$ son conjuntamente estacionarios en el sentido amplio. Los resultados para este caso son:

Las funciones media, potencia y autocorrelación de respuesta:

$$\bar{Y} = E[Y(t)] = \bar{X} \int_{-\infty}^{\infty} h(\xi) d\xi \quad (7.3)$$

$$\bar{Y}^2 = E[Y^2(t)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_{XX}(\xi_1 - \xi_2) h(\xi_1) h(\xi_2) d\xi_1 d\xi_2 \quad (7.4)$$

$$R_{YY}(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_{XX}(\tau + \xi_1 - \xi_2) h(\xi_1) h(\xi_2) d\xi_1 d\xi_2 \quad (7.5)$$

Las funciones de correlación cruzadas son

$$R_{XY}(\tau) = \int_{-\infty}^{\infty} R_{XX}(\tau - \xi) h(\xi) d\xi \quad (7.6)$$

$$R_{YX}(\tau) = \int_{-\infty}^{\infty} R_{XX}(\tau - \xi) h(-\xi) d\xi \quad (7.7)$$

Espectros de densidad de potencia

Sea un proceso aleatorio $X(t)$, que al menos es estacionario en el sentido amplio con espectro de densidad de potencia $\mathcal{L}_{XX}(\omega)$, aplicado a la entrada de una red lineal invariante en el tiempo que tiene una función de transferencia $H(\omega)$. El proceso aleatorio de respuesta $Y(t)$ tendrá un espectro de densidad de potencia.

$$\mathcal{L}_{YY}(\omega) = \mathcal{L}_{XX}(\omega) |H(\omega)|^2 \quad (7.8)$$

Los espectros de densidad de potencia cruzados están relacionados por $\mathcal{L}_{XX}(\omega)$ de la siguiente manera:

$$\mathcal{L}_{XY}(\omega) = \mathcal{L}_{XX}(\omega) H(\omega) \quad (7.9)$$

$$\mathcal{L}_{YX}(\omega) = \mathcal{L}_{XX}(\omega) H(-\omega) \quad (7.10)$$

8. PROCESOS ALEATORIOS PASA BANDAS

Asúmase nuevamente que $X(t)$ es un proceso aleatorio real, con media cero, estacionario en el sentido amplio y limitado en banda, que tiene un espectro de densidad de potencia definido por

$$\mathcal{L}_{XX} \begin{cases} \neq 0 & 0 < \omega_1 < |\omega| < \omega_1 + W_f \\ = 0 & \text{en otro caso} \end{cases} \quad (8.1)$$

Aquí ω_1 es constante y W_f es el espectro extendido del proceso. En la práctica, las señales aleatorias raramente están limitadas en banda, sin embargo, siempre hay frecuencias cuyo espectro de potencia es mayor y fuera de ellas es de poca importancia. Esto significa que la forma del espectro asumida es razonable para la mayoría de los problemas prácticos.

El proceso aleatorio $X(t)$ puede representarse de manera útil por

$$X(t) = X_c(t) \cos(\omega_0 t + \theta_0) - X_s(t) \sin(\omega_0 t + \theta_0) \quad (8.2)$$

donde ω_0 y θ_0 son constantes; $X_c(t)$ y $X_s(t)$ son procesos aleatorios pasa bajas con las siguientes propiedades

$$(1) X_c(t) \text{ y } X_s(t) \text{ son conjuntamente estacionarios en el sentido amplio} \quad (8.3a)$$

$$(2) E[X_c(t)] = 0, \quad E[X_s(t)] = 0 \quad (8.3b)$$

$$(3) E[X_c^2(t)] = E[X_s^2(t)] = E[X^2(t)] \quad (8.3c)$$

$$(4) R_{X_c X_s}(\tau) = R_{X_s X_c}(\tau) = \frac{1}{\pi} \int_0^{\infty} \mathcal{L}_{XX}(\omega) \cos[(\omega - \omega_0)\tau] d\omega \quad (8.3d)$$

$$(5) R_{X_c X_s}(\tau) = -R_{X_s X_c}(\tau) = \frac{-1}{\pi} \int_0^{\infty} \mathcal{L}_{XX}(\omega) \sin[(\omega - \omega_0)\tau] d\omega \quad (8.3e)$$

$$(6) \mathcal{L}_{X_c X_c}(\omega) = \mathcal{L}_{X_s X_s}(\omega) = L_p \left[\mathcal{L}_{XX}(\omega - \omega_0) + \mathcal{L}_{XX}(\omega + \omega_0) \right] \quad (8.3)$$

$$(7) \mathcal{L}_{X_c X_s}(\omega) = -\mathcal{L}_{X_s X_c}(\omega) = -jL_p \left[\mathcal{L}_{XX}(\omega - \omega_0) - \mathcal{L}_{XX}(\omega + \omega_0) \right] \quad (8.3g)$$

Aquí ω_0 es cualquier frecuencia conveniente en la banda $\omega_1 < \omega_0 < \omega_1 + W_f$ y $L_p[\cdot]$ representa tomar sólo la parte pasa bajas de la cantidad entre corchetes.

Una de las aplicaciones más útiles de la ecuación (8.2) es cuando $\mathcal{L}_{XX}(\omega)$ es simétrica alrededor de ω_0 . En este caso, $R_{X_c X_s}(\tau) = 0$, $R_{X_s X_c}(\tau) = 0$, $\mathcal{L}_{X_c X_s}(\omega) = 0$ y $\mathcal{L}_{X_s X_c}(\omega) = 0$

9. FILTROS ACOPLADOS

En un número de sistemas una señal determinística $f(t)$ se transmite y recibe junto con ruido $n(t)$ en el receptor. Muchas veces el problema en el receptor es decidir en qué instante de tiempo t_0 , si está presente la señal o sólo el ruido. La habilidad para tomar esta decisión mejora si la razón de la potencia de la señal al tiempo t_0 respecto a la potencia promedio del ruido es maximizada. De hecho es posible seleccionar un filtro, llamado filtro acoplado, que hará que la relación señal a ruido sea la máxima. En esta sección examinamos esta clase de filtros.

Caso de ruido coloreado

Considere algún filtro lineal invariante en el tiempo con respuesta al impulso y función de transferencia $h(t)$ y $H(\omega)$, respectivamente. La respuesta $f_o(t)$ generada mediante la aplicación de $f(t)$ a la entrada del filtro es

$$f_o(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) H(\omega) e^{j\omega t} d\omega \quad (9.1)$$

donde

$$f(t) \longleftrightarrow F(\omega) \quad (9.2)$$

La potencia promedio del ruido a la salida N_0 es

$$N_0 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathcal{L}_{NN}(\omega) |H(\omega)|^2 d\omega \quad (9.3)$$

donde $\mathcal{L}_{NN}(\omega)$ es el espectro de densidad de potencia del ruido $n(t)$ representado por el proceso aleatorio $N(t)$.

La razón de la potencia de la señal al tiempo t_0 respecto a la potencia promedio del ruido, que se quiere maximizar es

$$\left(\frac{S_0}{N_0} \right) = \frac{\left| \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) H(\omega) e^{j\omega t_0} d\omega \right) \right|^2}{\left(\frac{1}{2\pi} \int_{-\infty}^{\infty} \mathcal{L}_{NN}(\omega) |H(\omega)|^2 d\omega \right)} \quad (9.4)$$

Aplicando la desigualdad de Schwarz se tiene

$$\left(\frac{S_0}{N_0} \right) \leq \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|F(\omega)|^2}{\mathcal{L}_{NN}(\omega)} d\omega \right) \quad (9.5)$$

El máximo (igualdad) ocurre solamente cuando se selecciona la función de transferencia óptima como

$$H_{\text{opt}}(\omega) = \frac{1}{2\pi C} \frac{F^*(\omega)}{\mathcal{L}_{NN}(\omega)} e^{-j\omega t_0} \quad (9.6)$$

Caso de ruido blanco

En el caso del ruido blanco donde $\mathcal{L}_{NN}(\omega) = N_0/2$ (constante) para todas las frecuencias, la función de transferencia óptima para el filtro es

$$H_{\text{opt}}(\omega) = \frac{1}{\pi N_0 C} F^*(\omega) e^{-j\omega t_0} \quad (9.7)$$

Su respuesta al impulso se encuentra y es

$$h_{\text{opt}}(t) = \frac{1}{\pi N_0 C} f^*(t_0 - t) \quad (9.8)$$

Si $f(t)$ es real, su respuesta al impulso óptima es una réplica de $f(t)$ centrada en t_0 , pero atrasada en el tiempo.

ÍNDICE

<i>Capítulo I. Introducción</i>	7
I.1. Algunas definiciones	7
I.2. Un poco de historia	10
I.3. Ventajas de las redes digitales	13
I.3.1. Facilidad de diversos servicios	13
I.3.2. Facilidad de multiplexación	15
I.3.3. Facilidades para la utilización de tecnología moderna	15
I.3.4. Integración de la transmisión y la conmutación	17
I.3.5. Regeneración de señal	18
I.3.6. Funcionamiento con bajas relaciones señal-ruido	18
I.3.7. Facilidad de señalización	19
I.3.8. Facilidad criptológica	20
I.3.9. Facilidad de monitoreo	20
I.4. Desventajas de las redes digitales	21
I.4.1. Mayor requerimiento de ancho de banda	21
I.4.2. Necesidad de conversión A/D y D/A	21
I.4.3. Necesidad de sincronización de tiempo	22
I.4.4. Incompatibilidad con la red analógica existente	22
I.5. Conclusión	23
<i>Bibliografía</i>	23
 <i>Capítulo II. Teoría de la información</i>	25
II.1. Introducción	25
II.2. Medida de la información	27
II.3. Información mutua	27
II.4. Binit y bit	29
Ejemplo II.1.	30
Ejemplo II.2.	30
Ejemplo II.3.	30
Ejemplo II.4.	31
Ejemplo II.5.	31
II.5. Información promedio (entropía)	31
Ejemplo II.6.	32
II.6. Tasa de información	34
Ejemplo II.7.	34
Ejemplo II.8.	35
Ejemplo II.9.	35

Ejemplo II.10.
Ejemplo II.11.
Ejemplo II.12.
Ejemplo II.13.
Ejemplo II.14.
Ejemplo II.15.
Ejemplo II.16.
Ejemplo II.17.
Ejemplo II.18.
Ejemplo II.19.
Ejemplo II.20.
II.7. Información promedio en símbolos dependientes
II.8. Fuentes continuas
Ejemplo II.21.
Ejemplo II.22.
II.9. Canales de comunicación
II.10. Representaciones del canal (discreto)
Ejemplo II.23.
Ejemplo II.24.
Ejemplo II.25.
Ejemplo II.26.
II.11. Entropía condicional y conjunta
II.12. Capacidad del canal
Ejemplo II.27.
Ejemplo II.28.
Ejemplo II.29.
Ejemplo II.30.
Ejemplo II.31.
Ejemplo II.32.
II.13. Extensión de una fuente
Ejemplo II.33.
III.14. Códigos instantáneos y no instantáneos
Codificación de Shannon-Fano y Huffman
II.15. Canales continuos
II.16. El teorema de Shannon-Hartley
Ejemplo II.34.
Ejemplo II.35.
Ejemplo II.36.
II.17. Problemas
Bibliografía
Capítulo III. Muestreo
III.1. Algunas definiciones
III.1.1. Señales y fuentes analógicas
III.1.2. Señales y fuentes digitales
III.1.3. Clasificación de señales
III.1.4. Modulación
III.2. Diagrama a bloques de un sistema de comunicación digital

2.1. Subsistema de transmisión	77
2.2. Canal de comunicación	79
2.3. Subsistema de recepción	79
Muestreo ideal	79
Teorema de muestreo para señales pasa bajas	80
Traslapamiento (Aliasing)	87
Ejemplo III.1.	89
Ejemplo III.2.	90
Ejemplo III.3.	92
Ejemplo III.4.	93
Muestreo natural	94
Muestreo y retención	96
Multicanalización por división en el tiempo	100
4.7.1. Ancho de banda de TDM	100
Ejemplo III.5.	101
4.7.2. Sincronización	102
Interferencia entre pulsos (ISI)	102
Ejemplo III.6.	104
1. Espectro de coseno elevado	106
4.8.2. Espectro de seno elevado	108
4.8.3. Circuitos conformadores de pulsos	109
Ejemplo III.7.	110
Ejemplo III.8.	110
Ejemplo III.9.	111
Problemas	112
Bibliografía	115
 Título IV. Transmisión digital en banda base	117
1. Introducción	117
2. Modulación por pulsos codificados (PCM)	117
IV.2.1. El proceso de cuantificación	118
IV.2.2. El proceso de codificación	119
IV.2.3. Demodulación de señales PCM	119
3. Distorsión y niveles óptimos de cuantificación	121
Ejemplo IV.1.	123
Ejemplo IV.2.	126
Ejemplo IV.3.	128
Ejemplo IV.4.	129
Ejemplo IV.5.	130
Ejemplo IV.6.	130
4. Cuantificación no uniforme	131
IV.4.1. Compansión analógica	132
IV.4.2. Leyes de compansión	133
Ejemplo IV.7.	134
IV.4.3. Compansión digital	134
Ejemplo IV.8.	137
Ejemplo IV.9.	139
5. Jerarquías de sistemas PCM	140

IV.5.1. Portadora digital T1	141
IV.5.2. Portadora digital E1	144
IV.5.3. Multiplexaje digital de alto orden (UIT)	146
IV.6. Codificación de canal	152
IV.7. Probabilidad de error en transmisión binaria	157
IV.7.1. Niveles de decisión óptimos	162
IV.8. Técnicas de detección y corrección de errores	163
IV.9. Modulación delta	169
IV.9.1. Modulación delta adaptativa (ADPCM)	171
IV.10. Problemas	173
Bibliografía	177
<i>Capítulo V. Modulación digital</i>	179
V.1. Introducción	179
V.2. Eficiencia	179
V.3. Modulación binaria	180
V.3.1. Modulación en amplitud ASK	180
V.3.2. Modulación en frecuencia FSK	183
V.3.3. Un caso especial MSK (minimum shift keying)	186
V.3.4. Modulación en fase PSK	190
V.3.5. PSK Diferencial (DPSK)	193
V.4. Modulación multisímbolo	195
V.4.1. PSK multisímbolo M-PSK	196
V.4.2. Modulación por cuadratura de amplitud QAM	205
Ejemplo V.1.	209
Ejemplo V.2.	209
V.4.3. Esquemas FSK de M estados	209
V.5. Comportamiento en presencia de ruido	210
V.5.1. Filtro acoplado	211
V.5.2. Detección coherente de señales binarias	217
V.5.3. Detección diferencial coherente	221
V.5.4. Detección de señales multinivel en presencia de ruido	221
V.6. Recomendaciones internacionales para módems	223
V.7. Problemas	229
Bibliografía	230
<i>Apéndice A. Revisión de señales determinísticas y redes</i>	231
<i>Apéndice B. Revisión de teoría de señales aleatorias</i>	245

Principios de comunicaciones digitales
se terminó de imprimir en octubre de 1999
en los talleres de Sans Serif Editores, S.A. de C.V.,
Leonardo da Vinci 199, col. Mixcoac, 03910 México, D.F.
El tiro consta de 1 000 ejemplares más sobrantes
para reposición.
La composición tipográfica, la formación, la producción
y el cuidado editorial estuvieron a cargo
de Sans Serif Editores,
tels. 5611 37 37 y 5611 37 30.



Formato de Papeleta de Vencimiento

*El usuario se obliga a devolver este libro en la fecha
señalada en el sello mas reciente*

Código de barras. 2893843

FECHA DE DEVOLUCION

[illegible]

- Ordenar las fechas de vencimiento de manera vertical.
- Cancelar con el sello de "DEVUELTO" la fecha de vencimiento a la entrega del libro



2893843

Otros títulos en esta colección

Patrick Staelens, *El trabajo de los menores*

Luis Rodríguez, *El diseño preindustrial*

José Dolores Juárez Cervantes, *Sistemas de distribución de energía eléctrica*

Rafael Quintero, *Electrónica física*

Adalberto Cantú Chapa, *Análisis de diseño con diodos y transistores*

Mariem Henaine-Abed, *Planeación y control de la producción*

José Vega Luna y Gerardo Salgado Guzmán, *Prácticas de laboratorio de sistemas digitales*

Ana Lilia Laureano, *Programación orientada a objetos: un enfoque con tipos abstractos*

Juan González Márquez, *Introducción al derecho bancario mexicano*

Violeta Múgica y José de Jesús Figueroa, *Contaminación ambiental, causas y control*

Fernando Toledo Toledo, *Métodos computacionales para el análisis de sistemas de potencia*

Raymundo López Callejas, Juan Ramón Morales Gómez, Mabel Vaca Mier, Araceli Lara Valdivia y David Sandoval Cardoso, *Problemario de mecánica de fluidos*

Tomás David Navarrete González y José Ángel Rocha Martínez, *Colección de problemas resueltos para el curso de energías mecánica y eléctrica*

Rafael López Bracho, María Paula Ortuño Sánchez, Felipe Carrillo Romero y María Teresa Rodríguez Martínez, *Paquete computacional "Optimización en redes (versión 2.0) para Windows"*.



Principios de comunicaciones digitales está organizado de tal modo que constituye una firme base conceptual para los profesionales de las telecomunicaciones. La obra recoge la experiencia del autor en el proceso de enseñanza-aprendizaje e integra dudas, comentarios y aportaciones de alumnos de varias generaciones. Se encontrarán aquí reunidos temas fundamentales en el campo de los sistemas digitales de las telecomunicaciones, como la teoría de la información, el muestreo, la multicanalización por división de tiempo, la modulación por codificación de pulsos y la modulación digital, entre otros.